Report

The emergence of compositional communication in a synthetic ethology framework

Grant FA9550-06-1-0202

05 ID 012

José Fernando Fontanari
Instituto de Física de São Carlos, Universidade de São Paulo,
Caixa Postal 369
13560-970 São Carlos SP Brazil
Phone: +55-16-33739849, Fax: +55-16-33739877, e-mail:
fontanari@ifsc.usp.br

REPORT DOCUMENTATION PAGE

OMB No. 0704-0188

AFRL-SR-AR-TR-09-0253

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	
12 Aug 2005	Final Technical Report	01 Sep 05 – 31 Aug 08

4. TITLE AND SUBTITLE
FUNDS TO SUPPORT BASIC RESEARCH OF DR. JOSE FONTANARI WHOSE PROPOSALS ADDRESS EVOLUTION OF LANGUAGE TOGETHER WITH COGNITION.

5a. CONTRACT NUMBER

5b. GRANT NUMBER
FA9550-06-1-0202

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Jose Fontanari

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
INSTITUTO DE FISICA DE SAO CARLOS, UNIVERSIDADE DE SAO PAULO,
CAIXA POSTAL 369
13560 -970 SAO CARLOS SP BRAZIL
+55- 16-33739849, +55-16-33739877

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
USAF/IO
AFOSR
875 North Randolph Street
Arlington VA 22203

10. SPONSOR/MONITOR'S ACRONYM(S)
AFOSR/IO

11. SPONSORING/MONITORING AGENCY REPORT NUMBER
N/A

12. DISTRIBUTION AVAILABILITY STATEMENT

Distribution Statement A: Approved for public release. Distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The emergence of compositional communication in a synthetic ethology framework. Evolutionary language games have proved a useful tool to study the evolution of communication codes in communities of agents that interact among themselves by transmitting and interpreting a fixed repertoire of signals. Most studies have focused on the emergence of Saussurean codes (i.e., codes characterized by an arbitrary one-to-one correspondence between meanings and signals.) In this contribution, we argue that the standard evolutionary language game framework cannot explain the emergence of compositional codes – communication codes that preserve neighborhood relationships by mapping similar signals into similar meanings – even though use of codes would result in a much higher payoff in the case that signals are noisy. We introduce an alternative evolutionary setting in which the meanings are assimilated sequentially and show that the gradual building of the meaning-signal mapping leads to the emergence of mappings with the desired compositional property.

15. SUBJECT TERMS
COMPOSITIONAL COMMUNICATION, SYNTHETIC ETHOLOGY

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			J. Fillerup
Unclassified	Unclassified	Unclassified	Unclassified	11	19b. TELEPONE NUMBER (Include area code) (703)588.1781

Contents

1 Research activities

The main results of the research activities supported by the Air Force Office of Scientific Research (AFOSR) were described in great detail and made public in the five papers listed below and appended to the end of this report (see contents).

1. José F. Fontanari and Leonid I. Perlovsky, "*Evolving compositionality in evolutionary language games*", IEEE Transactions on Evolutionary Computation, published on-line doi:10.1109/TEVC.2007.892763
2. José F. Fontanari and Leonid I. Perlovsky, "*Inverse density dependence in the evolution of communication*", submitted to Journal of Theoretical Biology.
3. José F. Fontanari and Leonid I. Perlovsky, "*How communication can improve differentiation in the Modeling Field Theory framework*", Proceedings of the IEEE International Conference on Integration of Knowledge Intensive Multi-Agent Systems KIMAS07, Waltham, MA (ISBN: 1-4244-0945-4), pp. 151-156 (2007)
4. José F. Fontanari and Leonid I. Perlovsky, "*Language acquisition and category discrimination in the Modeling Field Theory framework*", Proceedings of the International Joint Conference on Neural Networks (IJCNN07), Orlando, FL.
5. Angelo Cangelosi, Vadim Tikhanoff, José F. Fontanari and Emmanouil Hourdakis, "*Integrating Language and Cognition: A Cognitive Robotics Approach*", invited contribution to IEEE Computational Intelligence Magazine.

The first two papers address the main topic of investigation of the research proposal. In particular, we have introduced a simple structured meaning-signal mapping, where meaning and signals are represented by integers and the metrics of the meaning and signal spaces are specified by the simple subtraction operation. Of particular relevance is our finding that structured (or compositional) communication codes cannot evolve within the traditional language evolutionary game setting: the evolutionary dynamics is plagued by local maxima that do not reflect the inner organization of the meaning and signal spaces. In the paper "*Evolving compositionality in evolutionary language games*" we have proposed an alternative learning scheme in which the individuals or agents learn the signal-meaning associations one by one – a procedure named sequential meaning assimilation. Provided the meanings are presented in an order that conforms to their proximity in the meaning space, this scheme works nicely and leads to the emergence of structured communication codes. In the paper "*Inverse density dependence in the evolution of communication*", we maintain the parallel or simultaneous presentation of all meanings but allow for some structure in the population, so that individuals adopting similar communication codes meet more frequently than individuals that adopt different codes. We then show that provided the aggregation pressure is sufficiently strong, structured codes are likely to emerge and become established in the population.

The last three papers do not address the emergence of compositionality issue directly. Rather, they focus on a more basic problem – the selective pressures responsible for the

evolution of communication. In particular, we show that the exchange of information between Modeling Field Categorization systems (or agents) can greatly improve the discriminating capability of each agent, in the sense they become capable of differentiating objects or categories that they could not distinguish without language. We note that an extension of paper 4 was accepted for publication in the special issue "Advances in Neural Networks Research - IJCNN 2007 Orlando" of the influential journal *Neural Networks*. Finally, paper 5 is a product of the research effort done in collaboration with the group lead by Dr. Angelo Cangelosi at the University of Plymouth, to implement the abstract framework described in papers 3 and 4 in a robotics scenario.

2 Acknowledgement of Sponsorship

Effort sponsored by the Air Force Office of Scientific Research, Air Force Material Command, USAF under grant number FA9550-06-1-0202. The U.S. Government is authorized to reproduce and distribute reprints for Government purpose notwithstanding any copyright notation thereon.

3 Disclaimer

The views and conclusions contained herein are those of the author and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the Air Force Office of Scientific Research or the U.S. Government.

4 Disclosure of inventions

I certify that there were no subject inventions to declare as defined in FAR 52.227-13, during the performance of this contract.

Evolving Compositionality in Evolutionary Language Games

José Fernando Fontanari and Leonid I. Perlovsky, *Senior Member, IEEE*

Abstract—**Evolutionary language games have proved a useful tool to study the evolution of communication codes in communities of agents that interact among themselves by transmitting and interpreting a fixed repertoire of signals. Most studies have focused on the emergence of Saussurean codes (i.e., codes characterized by an arbitrary one-to-one correspondence between meanings and signals). In this contribution, we argue that the standard evolutionary language game framework cannot explain the emergence of compositional codes—communication codes that preserve neighborhood relationships by mapping similar signals into similar meanings—even though use of those codes would result in a much higher payoff in the case that signals are noisy. We introduce an alternative evolutionary setting in which the meanings are assimilated sequentially and show that the gradual building of the meaning-signal mapping leads to the emergence of mappings with the desired compositional property.**

Index Terms—**Complexity theory, game theory, genetic algorithms, simulation.**

I. Introduction

THE CASE FOR the study of the evolution of communication within a multiagent framework was probably best made by Ferdinand de Saussure in his famous statement:

> "language is not complete in any speaker; it exists only within a collectivity... only by virtue of a sort of contract signed by members of a community" [1].

Translated into the biological jargon, this assertion means that language is not the property of an individual, but the extended phenotype of a population [2]. More than one decade ago, seminal computer simulations were carried out to demonstrate that cultural [3] as well as genetic [4] evolution could lead to the emergence of ideal communication codes (i.e., arbitrary one-to-one correspondences between objects or meanings and signals), termed Saussurean codes, in a population of interacting agents. Typically, the behavior pattern of the agents was modeled by (probabilistic) finite-state machines. The work by Hurford [3], in particular, set the basis of the Iterated Learning Model (ILM) for the cultural evolution of language, the typical realization of which consists of the interaction between two agents—a pupil that learns the language from a teacher [5]. In those studies, language is viewed as a mapping between meanings and signals. The communication codes that emerged from the agents' interactions are, in general, noncompositional or holistic communication codes, in which a signal stands for the meaning as a whole. In contrast, a compositional language is a mapping that preserves neighborhood relationships—similar signals are mapped into similar meanings. If there is a nontrivial structure in both meaning and signal spaces then, in certain circumstances, making explicit use of those structures may greatly improve the communication accuracy of the agents. The emergence of compositional languages in the ILM framework beginning from holistic ones in the presence of bottlenecks on cultural transmission was considered a breakthrough in the computational language evolution field [5]–[7]. The aim of this contribution is to understand how compositional communication codes can emerge in an evolutionary language game framework [3], [4], [8], [9].

The way we introduce the structure of the signal space (i.e., the notion of similarity between signals) into the rules of the language game is through errors in perception: the signals are assumed to be corrupted by noise so that they can be mistaken for one of their neighbors in signal space [8]. Similarly, the structure of the meaning space enters the game by rewarding the agents that prompted by a signal, infer a meaning close to the meaning actually intended by the emitter. Of course, the reward for incorrect but close inferences must be smaller than that granted for the correct inference of the intended meaning (see [9] for a similar approach). Hence, the role played by noise in this context is similar to the role of the bottleneck transmissions in the ILM framework, since both make advantageous the exploration of the detailed structure of the meaning-signal mapping. In particular, we show that once a Saussurean communication code is established in the population, i.e., all agents use the same code, it is impossible for a mutant to invade, even if the mutant uses a better code, say, a compositional one. This is essentially the Allee effect [10], [11] of population dynamics that asserts that intraspecific cooperation might lead to inverse density dependence, resulting in the extinction of some (social) animal species when their population size becomes small. Of course, this effect is germane to the outcome of biological invasions involving such species. We note that most realizations of the ILM circumvent this difficulty by assuming that the population is composed of two agents only, the teacher and the pupil, and that the latter always replaces the former. However, according to de Saussure (see quotation above), this is not an acceptable framework for

Manuscript received December 15, 2005; revised May 11, 2006 and January 2, 2007. This work was supported in part by the Air Force Office of Scientific Research, Air Force Material Command, USAF, under Grant FA9550-06-1-0202 and Grant FA8655-05-1-3031, and in part by CNPq and FAPESP under Project 04/06156-3. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purpose notwithstanding any copyright notation thereon.

J. F. Fontanari is with the Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, São Paulo 13560-970, Brazil (e-mail: fontanari@ifsc.usp.br).

L. I. Perlovsky is with Harvard University, Cambridge, MA 02138 USA and also with the Air Force Research Laboratory, Hanscom AFB, MA 01731 USA (e-mail: Leonid.Perlovsky@hanscom.af.mil).

Digital Object Identifier 10.1109/TEVC.2007.892763

language. In addition, a bias toward compositionality is built in the inference procedure used by the pupil to fill in the gaps due to transmission bottlenecks, in which some of the meanings are not taught to the pupil. This bias towards generalization, together with cultural evolution, seems to be the key ingredients to evolve compositionality in the ILM framework.

Understanding as well as demonstrating how innovations that increase the expressive power of individuals can spread through a population is the essence of any evolutionary explanation to language evolution [9]. Accordingly, the solution we propose to the problem of evolving a compositional code in a population of agents that exchange signals with each other and receive rewards at every successful communication event is the incremental assimilation of meanings, i.e., the agents construct their communication codes gradually, by seeking a consensus signal for a single meaning at a given moment. Only after a consensus is reached, a novel meaning is permitted to enter the game. This sequential procedure, which dovetails with the classic Darwinian explanation to the evolution of strongly coordinated system, allows for the emergence of fully compositional codes, an outcome that we argue is very unlikely, if not impossible, in the traditional language game scenario in which the consensus signals are sought simultaneously for the entire repertoire of meanings.

II. Model

Here, we take the more conservative viewpoint that language evolved from animal communication as a means of exchanging relevant information between individuals rather than as a byproduct of animal cognition or representation systems (see, e.g., [12] and [13] for the opposite viewpoint). In particular, we consider a population composed of N agents who make use of a repertoire of m signals to exchange information about n objects. Actually, since the groundbreaking work of de Saussure [1], it is known that signals refer to real-world objects only indirectly as first the sense perceptions are mapped onto a conceptual representation—the meaning—and then this conceptual representation is mapped onto a linguistic representation—the signal. Here, we simply ignore the object-meaning mapping (see, however, [14] and [15]) and use the words object and meaning interchangeably. To model the interaction between the agents, we borrow the language game framework proposed by Hurford [3] (see also [8]) and assume that each agent is endowed with separate mechanisms for transmission (i.e., communication) and for reception (i.e., interpretation). More pointedly, for each agent we define a $n \times m$ transmission matrix P whose entries p_{ij} yield the probability that object i is associated with signal j, and a $m \times n$ reception matrix Q the entries of which, q_{ji}, denote the probability that signal j is interpreted as object i. Henceforth, we refer to P and Q as the language matrices. In general, the entries of these two matrices can take on any value in the range [0,1] satisfying the constraints $\sum_{j=1}^{m} p_{ij} = 1$ and $\sum_{i=1}^{n} q_{ji} = 1$, in conformity with their probabilistic interpretation. In this contribution, however, we consider the case of binary matrices, in which the entries of Q and P can assume the values 0 and 1 only. There are two reasons for that. First, in the absence of errors in language learning, the evolutionary language game will eventually lead to binary transmission and reception matrices, regardless of the values of m and n, and of the initial choice for the entries of those matrices [16]. Therefore, our restriction of the entry values to binary quantities has no effect on the equilibrium solutions of the evolutionary game. In addition, these deterministic encoders and decoders were shown to perform better than their stochastic variants [17]. Second, by assuming that the transmission and reception matrices are binary, we recover the synthetic ethology framework proposed by MacLennan [4], a seminal agent-based work on the evolution of communication in a population of finite state machines (see also [18]).

Although the reception matrix Q is, in principle, independent of the transmission matrix P, results of early computer simulations have shown that in a noiseless environment, the optimal communication strategy is the Saussurean two-way arbitrary relationship between an object and a signal, i.e., the matrices P and Q are linked such that if $p_{ij} = 1$ for some object-signal pair i, j, then $q_{ji} = 1$ [3]. These matrices are associated to the Saussurean communication codes introduced before, provided there are no correlations between the different rows of the matrix P, i.e., the assignment object-signal is arbitrary.

A. The Evolutionary Language Game

Given the transmission and reception matrices, the communicative accuracy or overall payoff for communication between two agents, say I and J, is defined as [3], [8], [19]

$$F(I,J) = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{m} \left(p_{ij}^{(I)} q_{ji}^{(J)} + p_{ij}^{(J)} q_{ji}^{(I)} \right) \qquad (1)$$

from which we can observe the symmetry of the language game, i.e., both signaler and receiver are rewarded whenever a successful communication event takes place. By assuming such a symmetry, one ignores a serious hindrance to the evolution of language: passing useful information to another agent is an altruistic behavior [20], [21] that can be maintained in human societies thanks to the development of reciprocal altruism, in which unrelated individuals mutually benefit by exchanging the donor and the receiver roles multiple times [22]. However, the scarcity of empirical demonstrations of reciprocal altruism in nature, except for modern humans, motivated an alternative scenario for the evolution of language, namely, that human language evolved as a "mother tongue"—a communication system used among kin, especially between parents and their offspring [23].

In this contribution, we assume the validity of (1) and simply ignore the costs of honest signaling [20]. Hence, we take for granted the existence of special social conditions to foster reciprocal altruism among the agents or, alternatively, a mother tongue scenario in which the agents are related to each other. In this vein, it is interesting to note that although in the work by MacLennan [3] communication is defined following Burghardt [24] as "the phenomenon of one organism producing a signal that when responded to by another organism, confers some advantage to the signaler or his group" (see [25] for alternative definitions of communication), the actual implementation of the simulation rewards equally the two agents that take part in the successful communication event. In the case where only the receiver is rewarded, Saussurean communication fails to evolve [26].

Assuming, in addition, that each agent I interacts with every other agent $J = 1, \ldots, N$ $(J \neq I)$ in the population, we can immediately write down the total payoff received by I

$$F_I = \frac{1}{N-1} \sum_{J \neq I} F(I, J) \tag{2}$$

in which the sole purpose of the normalization factor is to eliminate the trivial dependence of the payoff measure on the population size N. Following the basic assumption of evolutionary game theory [27] this quantity is interpreted as the fitness of agent I. Explicitly, we assume that the probability that I contributes with an offspring to the next generation is given by the relative fitness

$$w_I = F_I \Big/ \sum_J F_J \tag{3}$$

which essentially implies that mastery of a public communication system adds to the reproductive potential of the agents [3].

There are several distinct ways to implement the language game. For instance, MacLennan [4] and Fontanari and Perlovsky [18] stick to the genetic algorithm approach (see, e.g., [28]) in which the offspring acquires both the transmission and reception matrices from its parent, assuming clonal or asexual reproduction. The offspring is identical to its parent except for the possibility of mutations that may alter a few rows of the language matrices. However, here we take a different viewpoint and reinterpret this genetic model within a learning context. We assume, in particular, that the offspring actually learns the language from its parent but that the learning is not perfect—there is a probability μ that the communication code it acquires is slightly different from its parent's. This very framework has been used to study the emergence of universal grammar and syntax in language [2], [29], [30].

An alternative learning scenario used by Nowak and Krakauer [8] assumes that the offspring adopt the language of its parent by sampling its response to every object k times. This approach makes sense only if the language matrices are not binary, though, as mentioned before, in the long run those matrices must become binary. For $k \to \infty$, the offspring is identical to its parent, which corresponds then to $\mu = 0$ in the previous learning scenario, whereas differences between parent and offspring arise in the case of finite $k > 1$. This sampling effect is qualitatively similar to the effect of learning errors in the scenario introduced before. For $k = 1$, already the first generation of offspring communicates through binary language matrices and so the sampling procedure is rendered ineffective. The reason is that a binary matrix P assigns each object to a unique signal (though this same signal can be used also for a distinct object), and so sampling the responses of the parent to the same object will always yield the same signal. As a result, the evolutionary process based on learning by sampling halts—the offspring become identical to their parents.

A similar but more culturally inclined approach is that followed by Hurford [3] and Nowak *et al.* [16]: instead of sampling the parent's responses, the offspring samples the responses of a certain number of agents in the population or even of the entire population. In this case, the hereditary component is lost since the offspring, in general, will not resemble its parent, and so natural selection has no say in the outcome of the dynamics. In the case of Hurford [3], there is still a strong genetic component as the offspring inherits from its parent its strategy of inference. Similarly, the ILM for the cultural evolution of language (see [5] and [7] for reviews) in its more popular version consists of two agents only, the teacher and the pupil who learns from the teacher through a sampling process identical to that just described. The pupil then replaces the teacher and a new, tabula rasa pupil is introduced in the scenario. This procedure is iterated until convergence is achieved. In this case, the payoff (2) plays no role at all in the language evolutionary process and the stationary language matrices will depend strongly on the inference procedure used by the pupil to create a meaning/signal mapping from the teacher responses. Of particular interest for our purpose is the finding that compositional codes emerge in the case that the learning strategy adopted by the pupil supports generalization and that this ability is favored by the introduction of transmission bottlenecks in the communication between teacher and pupil. Such a bottleneck occurs when the learner does not observe the signal for some objects. This contrasts with the sampling effect mentioned before in which the learner observes the signals to every object. In this contribution, we study whether and in what conditions compositional codes emerge in an evolutionary language game.

B. The Meaning-Signal Mapping

As already pointed out, language is viewed as a mapping between objects (or meanings) and signals and compositionality is a property of this mapping: a compositional language is a mapping that preserves neighborhood relationships, i.e., nearby meanings in the meaning space are likely to be associated to nearby signals in signal space [5]. At first sight, this notion looks contradictory to the well-established fact that the relation between a word (signal) and its meaning is utterly arbitrary. For instance, as pointed out by Pinker [31],

> "babies should not, and apparently do not, expect *cattle* to mean something similar to *battle*, or *singing* to be like *stinging*, or *coats* to resemble *goats*."

In fact, Pettito demonstrated that the arbitrariness of the relation between a sign and its meaning is deeply entrenched in the child's mind [32]. On the other hand, sentences like *John walked* and *Mary walked* have parts of their semantic representation in common (someone performed the same act in the past) and so the meaning of these sentences must be close in the meaning space. Since both sentences contain the word *walked* they must necessarily be close in signal space as well. Following Pinker, we acknowledge a significant degree of arbitrariness at the level of word-object pairing. This might be a consequence of a much earlier (prehuman) origin of this mechanism, as compared with seemingly distinctly human mind mechanisms for sentence-situation pairing. From a mathematical modeling perspective, however, such a distinction is not essential for our purposes, since the signals (sentences or words) can always be represented by a single symbol—only the "distance" between them will reflect the complex inner structure of the signal space. For instance, suppose there are only two words that we represent, without lack of generality by 0 and 1. Hence, a binary sequence

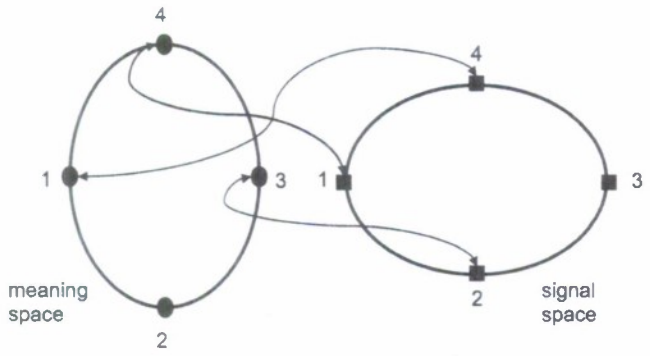

Fig. 1. Example of a mapping meaning-signal for $n = m = 4$. The integers here may be viewed as labels for complex entities (e.g., sentences). The large circles indicate cyclic boundary conditions so that, e.g., signal 1 is 1 unit distant from signals 2 and 4. The code represented in the figure has compositionality $C = 1$.

or, equivalently, its decimal representation can represent any sentence in this language. Here, the relevant distance between two such sentences is the Hamming distance rather than the result of the subtraction between their labeling integers. This notion, of course, generalizes trivially to the case when the sentences are composed of more than two types of words.

For simplicity, in this paper, we consider the case where both signals and meanings are represented by integer numbers and the relevant distance in both signal and meaning space is the result of the usual subtraction between integers. Fig. 1 illustrates one of the $n \times m$ possible meaning-signal mappings. A quantitative measure of the compositionality of a communication code is given by the degree to which the distances between all the possible pairs of meanings correlates with the distance between their corresponding pairs of signals [7]. Explicitly, let Δm_{ij} be the distance between meanings i and j, and Δs_{ij} the distance between the signals associated to these two meanings. Introducing the averages $\Delta\overline{m} = \sum_{(ij)} \Delta m_{ij}/p$ and $\Delta\overline{s} = \sum_{(ij)} \Delta s_{ij}/p$, where the sum is over all distinct pairs $p = n(n-1)/2$ of meanings, the compositionality of a code is defined as the Pearson correlation coefficient [7]

$$C = \frac{\sum_{(ij)}(\Delta m_{ij} - \Delta\overline{m})(\Delta s_{ij} - \Delta\overline{s})}{\left[\sum_{(ij)}(\Delta m_{ij} - \Delta\overline{m})^2 \sum_{(ij)}(\Delta s_{ij} - \Delta\overline{s})^2\right]^{1/2}} \quad (4)$$

so that $C \approx 1$ indicates a compositional code and $C \approx 0$ an unstructured or holistic code. This definition applies only to codes that implement a (not necessarily arbitrary) one-to-one correspondence between meaning and signal.

Strictly, here we do not address directly the emergence of compositionality, defined as the property that the meaning of a complex expression is determined by the meanings of its parts and the rules used to combine them. Rather, we focus on the emergence of structured communication codes, which preserve the topology of the meaning-signal mapping, in that similar meanings are associated with similar signals and *vice versa*. It seems that an important aspect of joint evolution of compositional cognition and compositional language is their evolution along with structural metric (or approximately metric) spaces of cognition and meaning. In this contribution, we assume that a metric space exists, and explore the consequences for the emergence of compositionality. The connection between structured and compositional meaning-signal mappings can be made explicit if we consider an artificial scenario for which there is a prescription to derive the meaning of the whole given the meaning of the elementary parts. (Such prescription is clearly ruled out in real language since context and previous knowledge play a crucial role in our understanding of any situation.) In this case, the distance between any two composite meanings could be inferred by comparing their components and, consequently, by introducing a metric in the meaning space.

Our approach ties in with the view that properties of language such as compositionality are emergent characteristics of the explosion of semantic complexity occurred during hominid evolution [33]. Semantic complexity means not only a large number of cognitive categories (meanings) but also an increase in their perceived interrelationships, which are inherent properties of the topology of the meaning space. In fact, the number of objects for which a person has separate words is not too large: a recent estimate suggests a vocabulary of around 60,000 base words for well-educated adult native speakers of English [34]. This is a not a very big number, and so it is reasonable to assume that object-word associations can be learned from examples, one by one. The number of situations that are combinations of objects, on the other hand, is larger than the number of all elementary particle events in the history of the Universe. This supports a need for the assumption of compositionality in language. As hinted in [33], a natural avenue to study the evolution of complex features of language (e.g., compositionality) is the increase of the complexity of the meaning space, which is exactly the approach we offer in this contribution.

C. Errors in Perception

So far as the communicative accuracy introduced in (1) is concerned, the structures of the meaning and signal spaces are irrelevant to the outcome of the evolutionary language game: the total population payoff is maximized when all agents adopt a code that implements a one-to-one correspondence between meanings and signals. Such a code is, of course, described by any one of the $n!$ permutation language matrices. The fact that ultimately all agents adopt the same communication code is a general result of population genetics related to the effect of genetic drift on a finite population [35]. To permit the structures of the meaning and signal spaces to play a role in the evolutionary game and so to break the symmetry among the permutation matrices so as to favor the compositional codes, we must introduce a new ingredient in the language game, namely, the possibility of errors in perception [8]. In fact, it is reasonable to assume that in the earlier stages of the evolution of communication the signals were likely to be noisy and so they could be easily mistaken for each other. The relevance of the structure of the signal space becomes apparent when we note that the closer two signals are, the higher the chances that they are mistaken for each other. This aspect of the model can be described by an agent-independent $m \times m$ confusion matrix E, the entries of which e_{ij} yield the probability of signal j being observed as signal i due to corruption by noise [8], [9].

To introduce the structure of the meaning space in the language game, we note first that (1) has a simple interpretation in the case of binary, but not necessarily permutation, language matrices: both signaler and receiver are rewarded with 1/2 unity of payoff whenever the receiver interprets correctly the meaning of the emitted signal. Otherwise, there is no reward to any of the two parts, no matter how close the inferred meaning is from the correct one. This gives us a clue as to how to modify the model in order to take into account the meaning structure—just ascribe some small reward value to both agents if the inferred meaning is close to the intended one. In fact, giving value to decisions which are not the best ones is a common assumption in decision and game theory [36] and seems to be consistent with what is actually observed in nature since, clearly, not every misinterpretation is equally harmful [9]. Consider for instance the Vervet monkey alarm calls [37]: misinterpreting a snake alarm for a leopard one, and hence running to a tree instead of standing up and looking in the grass, is clearly much better than misinterpreting it for an eagle call.

Following Nowak *et al.* [8] and Zuidema [9], we can formalize the notion of meaning similarity by introducing another agent-independent matrix, the $n \times n$ value matrix V, so that v_{ij} yields the payoff attributed to an agent which infers meaning i when the actual meaning the signaler intended to transmit was j. Hence, the overall payoff for communication between agents I and J becomes [9]

$$F(I,J) = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} v_{ij}\left[P^{(I)} \times \left(E \times Q^{(J)}\right) + P^{(J)} \times \left(E \times Q^{(I)}\right)\right]_{ij} \quad (5)$$

where $\times$ stands for the usual matrix multiplication. Note that (1) is recovered in the case that both value and confusion matrices are diagonal.

In particular, here we will consider the simple case in which there is a nonzero probability $\varepsilon \in [0,1]$ that a signal, say signal j, be mistaken for one of its nearest neighbors only, $e_{j-1,j} = \varepsilon/2$ and $e_{j+1,j} = \varepsilon/2$. Of course, the probability that a signal is not corrupted by noise is $e_{j,j} = 1 - \varepsilon$. If signal j is in the boundary, $j = 1$ or $j = m$, then we use the cyclic structure of the signal space to set $e_{0,1} = e_{m,1} = \varepsilon/2$ and $e_{m+1,m} = e_{1,m} = \varepsilon/2$. So, in the example of Fig. 1, signal 4 can be mistaken only for signals 3 or 1 with probability ε. Similarly, agents are rewarded only if the inferred meaning is one of the nearest neighbors of the intended meaning. For example, if the intended meaning is j, then the only nonzero entries of the value matrix V are $v_{j,j} = 1$, $v_{j+1,j} = r$, and $v_{j-1,j} = r$. Meanings in the boundary, $j = 1$ and $j = n$, are treated using the cyclic boundary conditions as explained for the signal space. Here, $r \in [0,1]$ is a parameter that measures the advantage, in terms of payoff, of using a compositional communication code rather than a Saussurean one.

Together with the presence of noise, this last ingredient—nonzero reward for inferring a meaning close to the correct one—should favor, in principle, the emergence of compositional communication codes in an evolutionary game guided by Darwinian rules. In what follows, we will show that the problem of evolving efficient communication codes within an evolutionary framework, whether in the presence or not of noise, is more difficult than previously realized [4], [16], [18]. This problem differs from usual optimization problems tackled with evolutionary algorithms in that the maximization of the average population payoff requires a somewhat coordinated action of the agents. It is of no value for an agent to exhibit the correct "genome" (i.e., the transmission and reception matrices) if it cannot communicate efficiently with the other agents in the population because they use different language matrices.

The emergent view of compositionality adopted here differs from the approach followed by Nowak *et al.* [29] to study the evolution of syntactic (or combinatorial) communication. In that work, the conditions at which syntax is advantageous over nonsyntactic or holistic languages were determined, namely, when the number of required signals to express the relevant meanings exceeds some threshold value. (It should be noted that combinatorial communication has its disadvantages too, since it boosts the potential for deception [38].) However, the finding that the adoption of a particular communication code is better for the population, in that it yields a higher overall payoff, is no guarantee that such code will actually spread in the population. On the contrary, in this contribution we show that the Allee effect will prevent its spreading. Additional assumptions, such as the semantic continuity of incremental learning proposed here, seem to be necessary to guarantee the emergence of compositional codes.

III. Population Dynamics

We assume that the offspring learn their languages from their parents. Were it not for the effect of errors during learning, which results in small changes in the language matrices, the offspring would be identical to their parents. Like mutations in the genetic setup, these learning errors allow for the variability of the agents, and thus for the action of natural selection.

We start with N agents (typically $N = 100$) whose binary language matrices are set randomly. Explicitly, for each agent and for each meaning $i = 1, \ldots, n$, we choose randomly an integer $j \in \{1, \ldots, m\}$ and set $p_{ij} = 1$ and $p_{ik} = 0$ for $k \neq j$. Similarly, for each signal $j = 1, \ldots, m$, we choose an integer $i \in \{1, \ldots, n\}$ and set $q_{ji} = 1$ and $q_{jk} = 0$ for $k \neq i$. This procedure guarantees that initially P and Q are independent random probability matrices. Note that, in general, they are not permutation matrices at this stage. To calculate the total payoff of a given agent, say agent I, we let it interact with every other agent in the population. At each interaction, the emitted signal can be mistaken for one of the neighboring signals with probability ε. According to (5), at each communication event (an interaction) agent I receives the payoff value 1/2 if the receiver guesses the intended meaning of the signal that I has emitted, the payoff value $r/2$ if the receiver guessing is one of the nearest neighbors of the intended meaning, and payoff value 0, otherwise. Of course, the receiver obtains the same payoff accrued to agent I. Once the payoffs or fitness of all N agents are tabulated, the relative payoffs can be calculated according to (3), and then used to select the agent that will contribute with one offspring to the next generation.

To keep the population size constant, we must eliminate one agent from the population. To do that we will use two strategies: 1) to choose the agent to be eliminated at random, regardless of its fitness value and 2) to use an elitist strategy which eliminates the agent with the lowest fitness value. In both cases, the recently produced offspring is spared from demise. The first selection procedure is Moran's model of population genetics [35]. Both procedures differ from the standard genetic algorithm implementation [28] in that they allow for the overlapping of generations, a crucial prerequisite for cultural evolution which may be relevant when learning is allowed. In practice, however, Moran's model does not differ from the parallel implementation in which the entire generation of parents is replaced by that of the offspring in a single generation. We define the generation time t as the number of generations needed to produce N offspring with the consequent elimination of the same number of agents.

Finally, to allow for the appearance of novel codes (or language matrices) in the population, changes are performed independently on the transmission and reception matrices of the offspring with probability $u \in [0, 1]$. Explicitly, the transmission matrix P is modified by changing randomly the signal associated to an also randomly chosen meaning with probability u. A similar procedure updates the reception matrix Q. Hence, the probability that the same offspring has its transmission and reception matrices simultaneously altered by errors is u^2 and the probability that it will differ somehow from its parent is $\mu = 1 - (1 - u)^2$. Henceforth, we will refer to μ as the probability of error in language acquisition.

To facilitate comparison between different evolutionary algorithms, we define a properly normalized average payoff of the population

$$G = \frac{1}{nN} \sum_{I=1}^{N} F_I \tag{6}$$

so that $G \in [0, 1]$. The maximum value $G = 1$ is reached for Saussurean codes in the case of noiseless communication.

In Fig. 2, we present the effect of the inaccuracy in language acquisition on the average payoff of the population for the simplest situation, namely, $\varepsilon = 0$ (the receiver always gets the original signal) and $r = 0$ (only inference of the correct meaning is rewarded). The results show a stark difference between the elitist and the usual evolutionary strategy regarding the form they are affected by learning errors. Whereas the performance of Moran's model is degraded for high error rates [39], reaching the payoff of random binary matrices for $\mu = 1$, the elitist strategy actually benefits from those errors and gets to the maximum payoff for the highest possible error rate. In fact, for small but nonzero values of the error rate, the communication accuracy of the elitist strategy is practically constant and starts to increases only after μ crosses some threshold value $\mu \approx 0.02$. The performance of Moran's model, on the other hand, indicates the existence of an optimum value of the learning error for which the communication accuracy is maximum. Longer runs do not show any significant change of the pattern illustrated in Fig. 2. What enables the elitist strategy to take advantage of errors is the overlapping of generations together with the immediate removal

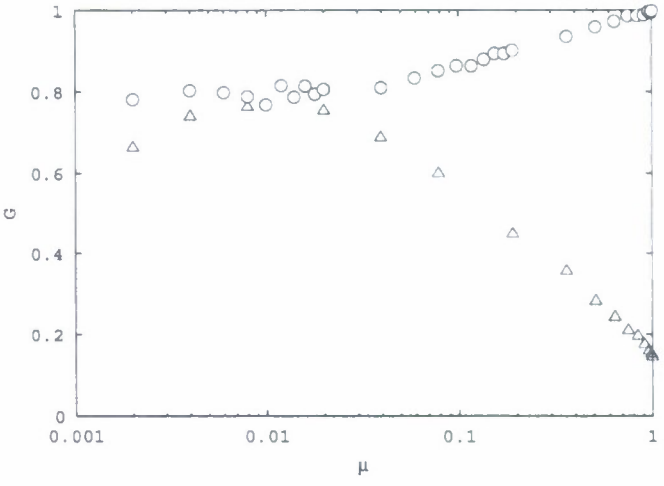

Fig. 2. Normalized average payoff G of the population as function of the probability of error in language acquisition μ in the case of $N = 100$ agents communicating about $n = 10$ meanings using $m = 10$ signals. The evolution was followed until $t = 2 \times 10^3$ for the elitist strategy (o) and until $t = 10^4$ for Moran's model (Δ). The symbols represent the average of over 50 independent runs. The error bars are smaller than the symbol sizes. For $\mu = 0$, we find $G = 0.255 \pm 0.005$ for both strategies, whereas for random language matrices we find $G = 0.1 \pm 0.0001$. The other parameters are $\varepsilon = r = 0$. The search space is the $m^n \times n^m$ space spanned by the two independent binary probability matrices P and Q.

of unfit agents from the population. This combination prevents the accumulation of inefficient agents in the population and the consequent degradation of the communication performance observed in Moran's model. Moreover, by eliminating the agent that performs worse in the language game, the elitist strategy adds an extra kick to the selective pressure towards better communication codes, in addition to the fitness regulation of offspring production described in (3).

The reason the elitist strategy can guide the population to a regime of practically perfect communication accuracy even in the presence of a constant flux of inefficient mutants ($\mu = 1$) is that a defective offspring, though spared from demise at birth, will almost certainly be purged from the population in the next step. We recall that a single generation comprises N such generation/elimination steps. In this scheme, the population can maintain at most a single defective agent, thus resulting in a reduction of the maximum normalized payoff by a factor $1/nN$. In view of the remarkable effectiveness of the elitist strategy to maximize the communication accuracy of the population, in what follows we will present the results for that strategy only.

Fig. 3 presents the average communication accuracy for 100 independent runs (populations) in a generic case in which the parameters ε and r, which couple the dynamics with the distances in the signal and meaning spaces are nonzero. Now, since the communication between any two agents is affected by noise, we must adopt a slightly different procedure to evaluate the payoff of the entire population. As before, we follow the evolutionary dynamics (i.e., the differential reproduction and learning-with-error procedures) until $t = 2 \times 10^3$, then we store the language matrices of all N agents. Keeping these matrices fixed, we evaluate the average population payoff in 100 contests. A contest is defined by the interaction between all pairs of agents in the population. Actually, according to (5),

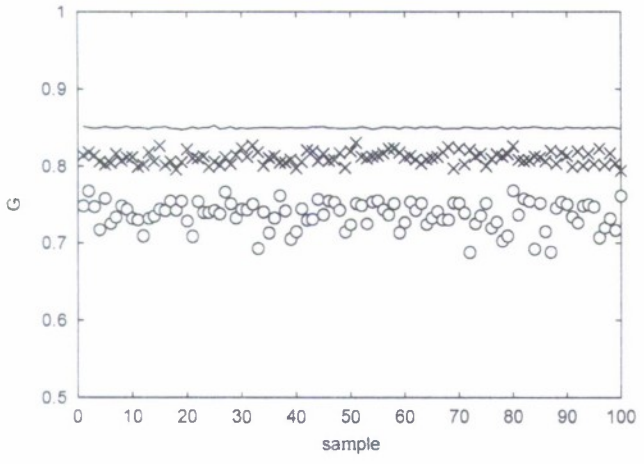

Fig. 3. Normalized average payoff for the elitist (o) strategy at $t = 2 \times 10^3$ for 100 independent sample runs of the evolutionary dynamics. These results are compared with that of a fully compositional code (solid line) and of Saussurean codes (×). The parameters and search space are the same as in Fig. 2 with $\mu = 1$, except that now we have included a pressure for compositionality: the signals are corrupted with probability $\varepsilon = 0.2$ and the ratio between the payoffs for inferring a close and the correct meaning is $r = 0.25$. The optimal, compositional code yields $G \approx 0.85$ and the typical payoff of a Saussurean code is $G \approx 0.80$.

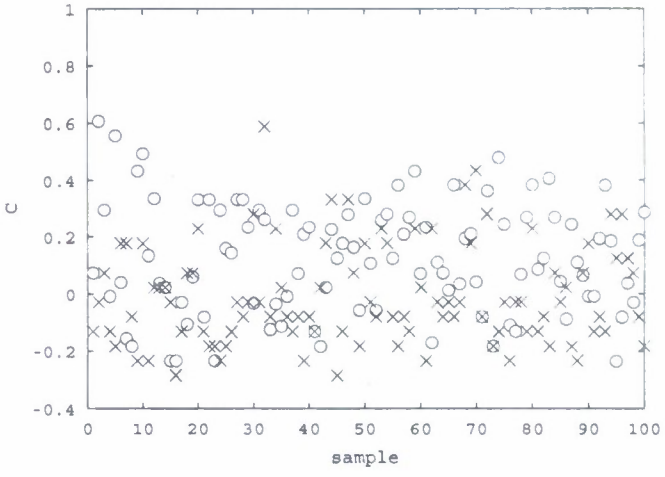

Fig. 4. Compositionality of the code carried by the agent with the highest payoff in the runs shown in Fig. 3. The compositionality of the perfect compositional code is, by definition, $C = 1$. There is a slight tendency to compositionality in the codes produced by the elitist (o) strategy as compared with those of the Saussurean codes (×).

each interaction comprises two communication attempts, since any given agent first plays the role of the emitter and then of the receptor. Hence, a contest amounts to $N(N-1)$ communication events. Of course, in the noiseless case ($\varepsilon = 0$), the payoff obtained would be the same in all contests. The procedural changes are needed to average out the effects of noise. For instance, in a single interaction two perfectly compositional codes could perform worse than two holistic codes if, by sheer chance, the signals happen to be corrupted only during the interaction of the compositional codes. To avoid such spurious effects the payoffs resulting from the interactions between any two agents are averaged out over 100 different interactions.

For the purpose of comparison, in Fig. 3 we also present the results for a population of agents carrying the same perfectly compositional code ($C = 1$), as well as for a similarly homogenous population of agents carrying identical Saussurean codes. These are control populations that in contrast to the elitist populations, do not evolve. In the absence of noise, these control populations would reach the maximum allowed payoff, $G = 1$. We note that a perfectly compositional code is not a Saussurean code, in the sense that the one-to-one mapping between meaning and signals is not arbitrary. The elitist strategy seems to face great difficulties even to find a Saussurean code, as compared with the performance in the noiseless case (see Fig. 2) for instance, not to mention to find the optimum, perfect compositional code. Actually, in the presence of noise, the performance of the Saussurean code seems to pose an upper limit to the performance of the elitist strategy by acting as an attractor to the evolutionary dynamics.

It is instructive to calculate the average payoff G_c of a population composed of identical agents carrying a perfect compositional code. Consider the average payoff received by a given agent, say I, in a very large number of interactions with one of its siblings, say J. When I plays the signaler role its average payoff is $(1-\varepsilon) \times 1/2 + \varepsilon \times r/2$, which, by symmetry, happens to be the same average payoff I receives when it plays the receiver role. Since all agents are identical, the expected payoff of any agent equals that of the population. Hence

$$G_c = 1 - \varepsilon(1 - r). \tag{7}$$

We can repeat this very same reasoning to derive the average payoff G_S of a homogenous population of Saussurean codes. In this case, by playing the signaler, I receives the average payoff $(1-\varepsilon) \times 1/2 + \varepsilon \times 2/(n-1) \times r/2$, where the factor $2/(n-1)$ accounts for the fact that the reward $r/2$ is obtained only if the inferred meaning is one of the two neighbors of the correct meaning. This reasoning is valid for $n > 2$ only, since for $n = 2$ each meaning has a single neighbor, and so there is no difference between Saussurean and compositional codes. Taking into account the payoff received by I when playing the receiver yields

$$G_S = 1 - \varepsilon + \frac{2\varepsilon}{n-1} r \tag{8}$$

for $n > 2$. Note that $G_c > G_S$ for $n > 3$. Similarly to the case $n = 2$, the Saussurean codes for $n = 3$ are compositional codes because of the cyclic boundary conditions in the meaning space. In Fig. 4, we show the compositionality of the code carried by the agent with the largest payoff value in each of the runs used to generate the data of Fig. 3. Although there is a slight tendency to compositionality in the codes produced by the elitist strategy, it is fair to say that the pressure to generate compositional code has not worked as expected, despite the clear advantage of such codes given the conditions of the experiment (see Fig. 3). As pointed out, the reason for that might be that the Saussurean codes act as barriers (local maxima) from which the evolutionary dynamics cannot escape, thus impeding it from reaching a perfect compositional code (global maximum).

The results depicted in Fig. 3 expose clearly the failure of the language evolutionary framework to produce efficient communication codes when the receiver must interpret noisy signals. To rule out the possibility that the cause of such failure

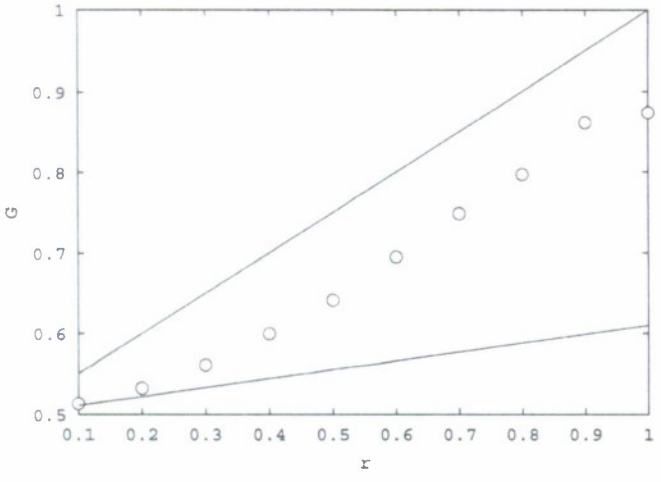

Fig. 5. Average payoff resulting from 100 independent runs of the noisy evolutionary language game with the search space restricted to permutation matrices (o) as a function of the pressure for compositionality. The error bars are smaller than the symbol sizes. The upper straight line is the function $G_c = (1 + r)/2$ that yields the average payoff of a perfect compositional code and the lower straight line is $G_S = 0.5 + 0.11r$ that yields the average payoff of a Saussurean code (see (7) and (8)). The parameters are $\varepsilon = 0.5, \mu = 0.9, N = 100$, and $n = m = 10$.

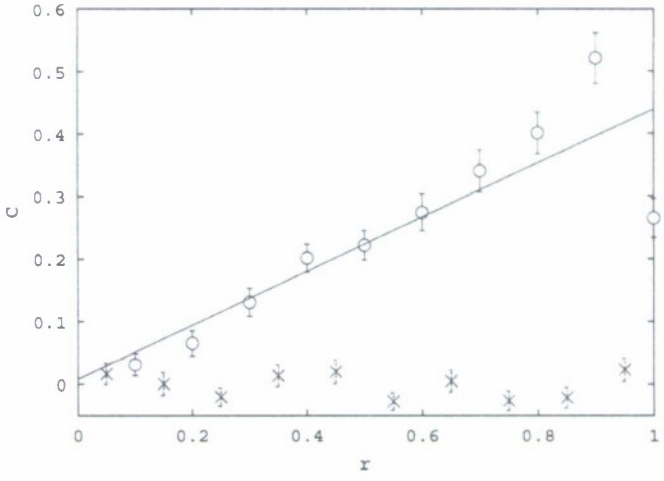

Fig. 6. Average compositionality of the 100 evolved communication codes (o) whose payoffs are exhibited in Fig. 5, as well as of the same number of Saussurean codes (×). The compositionality of a perfect compositional code is $C = 1$ by definition. The linear fitting of the average compositionality of the evolved codes yields a slope of ≈0.43.

was the initial unlikely decoupling between production and interpretation, in the following, we will restrict the search space to that of Saussurean codes. Hence, for any agent, the transmission matrix P is a permutation matrix and the reception matrix Q has entries given by $q_{ji} = 1$ if $p_{ij} = 1$ and 0 otherwise (Q is also a permutation matrix). The initial population is composed of N agents adopting distinct Saussurean codes. To guarantee that all new codes generated by mutations stay within our search space, we modify the mutation procedure so that with probability μ the signal associated to a randomly chosen meaning, say i, is exchanged with the signal associated to another randomly chosen meaning, say k. This corresponds to the interchange of the rows i and k of the transmission matrix. The reception matrix is then updated accordingly. The sole genetic strategy we use in the forthcoming simulations is the elitist one, in which the worst performing agent is replaced by the offspring of the agent chosen by rolling the fitness wheel.

In Fig. 5, we show the results of the experiments with the evolutionary search restricted to the space of permutation matrices. The procedure we use here was the same as that employed to draw Figs. 3 and 4: after the evolutionary dynamics has settled to an equilibrium (i.e., all agents are using the same communication code, except for single temporary mutants), the resulting homogeneous population is then left to interact for 100 contests and the average payoff is recorded. However, instead of exhibiting the payoff obtained in the 100 independent runs as in Fig. 3, we exhibit in Fig. 5 only the average payoff calculated over those runs. Hence, to obtain each data point of this figure we need to generate a set of data similar to that used to draw Fig. 3. We choose as the independent variable the ratio between the payoffs for inferring a neighbor of the correct meaning and the correct meaning (r), which can be interpreted also as a selective pressure for evolving compositional codes. For the sake of comparison, Fig. 5 also shows the average payoffs of perfect compositional and random Saussurean codes.

The results in Fig. 5 indicate that for $r = 0$, the performance of the communication codes, regardless of whether random, compositional or evolved, are identical. Explicitly, in this case, we find $G = 1 - \varepsilon$ for any one-to-one mapping. Since the search space is now restricted to the space of permutation matrices, it is not a surprise that the payoffs of the Saussurean codes serve as lower bounds to those of the evolved codes. This trivial finding should not be confused with the unexpected result exhibited in Fig. 3, that the payoffs of the Saussurean codes serve as upper bounds to the payoffs of the evolved codes when the search space is enlarged to cover all binary language matrices. The results in Fig. 5 show clearly that, despite the fact that compositionality can greatly improve the communication payoff of the population (see upper straight line in that figure), the evolved codes fall short of taking full advantage of the structure of the meaning-signal space to cope with the noise in the communication. As a result, the evolved codes are far from the optimal, perfect compositional codes, although they fare better than the Saussurean codes. Fig. 6 explains the reason for that: the evolutionary dynamics actually succeeded to produce partially compositional codes, thus reducing the deleterious effects of noise.

It is interesting that the payoffs of the Saussurean codes increase when the pressure for compositionality increases [see Fig. 5 and (8)], although they remain largely noncompositional in average (see Fig. 6). The key to the explanation of this result is found in Fig. 4, where we can see that half of the samples of the random Saussurean codes exhibit a positive value of the compositionality, which is then associated to a payoff value greater than $1 - \varepsilon$ ($= 0.8$ in that case), while the representatives of the other half have a payoff of $1 - \varepsilon$ at worst. It is clear that the resulting average payoff must be an increasing function of r.

The reason that the evolutionary dynamics failed to produce perfect compositional codes, despite their obvious advantage to

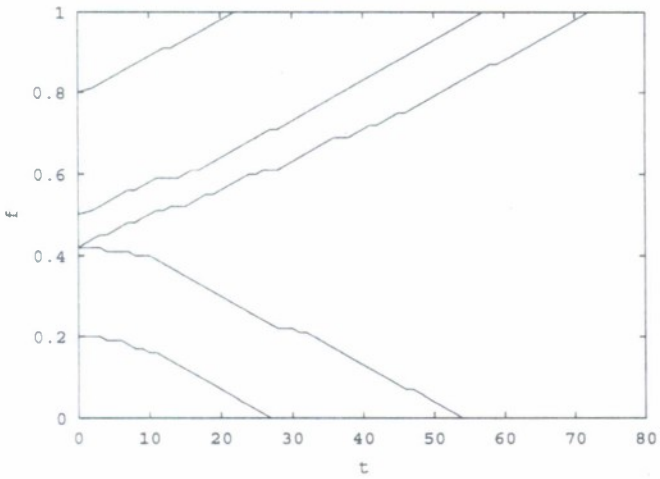

Fig. 7. The evolution of the fraction f of agents carrying a perfect compositional code in an experiment in which they compete against agents carrying a Saussurean code. The parameters are $\varepsilon = 0.5$, $r = 0.25$, $N = 100$, and $n = m = 10$. The initial population is set so that (from top to bottom) $f = 0.8$, 0.5, 0.42, 0.419, and 0.2.

cope with noisy signals, is that once a nonoptimal communication code has become fixed (or even almost fixed) in the population, mutants carrying better codes cannot invade. In fact, those mutants will most certainly do badly when communicating with the resident agents and, as a result, will quickly be removed from the population. As pointed out, this is essentially the Allee effect of population dynamics.

The task faced by the evolutionary algorithm here is of an essentially different nature from that tackled in typical optimization problems in which the fitness of an agent is frequency independent. In such a case, a fitter mutant can always invade the resident population. To stress this phenomenon, Fig. 7 illustrates the competition between a fraction f of agents carrying (the same) perfect compositional code and a fraction $1 - f$ of agents carrying (the same) Saussurean code. This simulation is implemented using the elitist procedure described before, except that learning errors are not allowed, so that at any time an agent can carry only one of the two types of codes set initially. Alternatively, Fig. 7 can be interpreted as the competition between two different strategies: the perfect compositional and the holistic strategies. We can easily estimate the minimum fraction f_m of perfect compositional codes above which this strategy dominates the population. It is simply

$$\frac{f_m}{1 - f_m} = \frac{G_S}{G_c} \tag{9}$$

with G_c and G_S given by (7) and (8), respectively. For the parameters of Fig. 8, this estimate yields $f_m \approx 0.46$, which within statistical errors, is in very good agreement with the single run experiment described in the figure. Repetition of this experiment using Moran's model rather than the elitist strategy leads to the same result, except that the fixation of the winner strategy takes much longer—about 100 times longer than the fixation times exhibited in Fig. 7.

This simple analysis of the competition between suboptimal Saussurean codes and the optimal compositional codes lends support to our previous conclusion that compositional codes do not evolve within the usual language evolutionary game framework because the evolutionary dynamics is very likely to get trapped in the local maxima—the Saussurean codes.

IV. Incremental Meaning Assimilation

What we have been trying to do up to now is to evolve in a single shot a communication code that associates each of the n meanings (or objects) to one of the m signals available in the repertoire of the agents. As pointed out, in the case that the meaning-signal mapping has a nontrivial underlying structure, the optimal association is not completely arbitrary in the sense that in the presence of noise some codes (i.e., the perfect compositional codes) result in a much better communication accuracy than codes that implement an arbitrary one-to-one correspondence between meaning and signals (Saussurean codes). The results of the previous simulations lead us to conclude that it is very unlikely, if not impossible, that evolution through natural selection alone could take advantage of the structure of the meaning-signal space to produce the optimal, perfect compositional codes.

The outcome would be very different, however, if the task posed to the population were to reach a consensus on the signals to be assigned to the meanings in a sequential manner. In other words, let us consider the situation in which each agent has m signals available (here we set $m = 10$) and the population needs to communicate about a single meaning, say $i = 1$. The search space is reduced then to the space of the $1 \times m$ permutation matrices. (We restrict the search space to that of permutation matrices, for simplicity.) Once the consensus is reached (i.e., the signal assigned to meaning $i = 1$ is fixed in the population), a new meaning is presented and the population is then challenged to find a consensus signal for that meaning. The procedure is repeated until each of the $n = m$ meanings are associated to a unique signal.

In the case of structured meaning-signal mappings, the order of presentation of meanings to the population plays a crucial role on the outcome of this strategy, which we term sequential meaning assimilation. In particular, success is guaranteed only if the novel meaning is a neighbor of the previously presented meaning (e.g., $i = 2$ or $i = N$ in the case the previous assimilated meaning was $i = 1$). In this case, the question is whether the population will reach a consensus on a signal that is also a neighbor of the signal assigned to the previous meaning. Curve (a) of Fig. 8 shows that this scheme works neatly, and yields a fully compositional code provided that $\varepsilon \neq 0$ and $r \neq 0$. We note that when the number of assimilated meanings is less than the size of the repertoire of signals m, the payoff of the sequential assimilation scheme [curve (a)] falls below the average payoff a fully compositional code (dashed horizontal line), because until all meanings are presented, the codes produced by that scheme cannot take full advantage of the topology of the meaning and signal spaces. The following example explains the reason this is so. Consider the situation in which two meanings were assimilated, say $i = 1,2$ and the signals assigned to them were $j = 6,7$, respectively. The agents will receive no reward if the corrupted signals become 5 or 8 (we recall that $m = 10$ in this experiment), since at this point there are no meanings associated to these altered signals. In contrast, reward is always

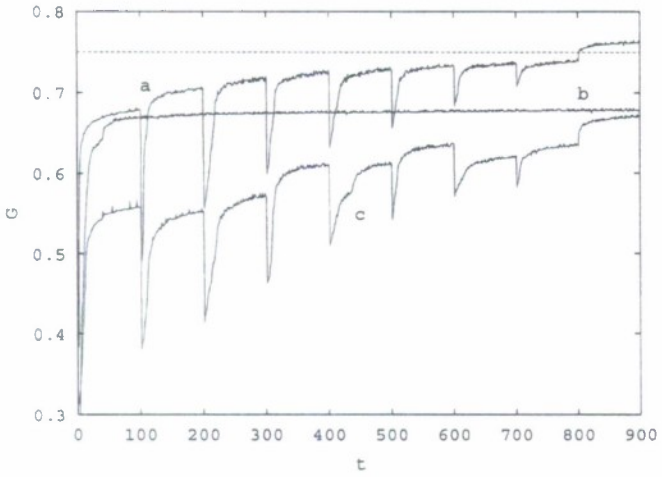

Fig. 8. Average payoff of the population when the task is to produce consensus signals to n meanings presented sequentially at the time intervals $\Delta t = 100$. In curve (a), the new meaning is a neighbor of the previous one, whereas in curve (c), the order of presentation of the meanings is random. The result for the usual batch algorithm, in which all meanings are presented simultaneously, is shown in curve (b). The dashed horizontal line indicates the average performance of perfect compositional codes. The parameters are $\varepsilon = 0.5$, $r = 0.5$, $N = 100$, and $n = m = 10$.

guaranteed for the fully formed compositional code since, by definition, all meanings are assimilated at the very outset in this case. Of course, as seen in Fig. 8, this "surface" effect is attenuated as more meanings are assimilated. The fact that the final payoff of the single run displayed in curve (a) ends up being greater than the (theoretical) average payoff of the perfect compositional code is simply a statistical fluctuation. Curve (c) in Fig. 8 illustrates the failure of the sequential presentation scheme when the order of presentation of meanings is random. In fact, if the meanings are presented in an arbitrary order, say $i = 3$ after $i = 1$, then there is no selection pressure to prevent that the signal assigned to $i = 3$ be one of the neighbors of the signal associated to $i = 1$. Eventually, when the meaning $i = 2$ is presented this optimal signal will be unavailable to the agents, precluding thus the emergence of a compositional code. Finally, we note that the incremental learning scheme would work all the same if the repertoire of meanings were left fixed and the signals were presented one by one.

The proposed solution to the evolution of compositional codes in an evolutionary language game framework could be questioned, because it relies on the assumption that the new meanings entering the population repertoire must be closely related to the already assimilated meanings. However, this seems to be the manner in which the perceptual systems work during categorization: new meanings are usually hierarchically related to the assimilated ones and this could be, for instance, the reason for Zipf's law of languages [40], [41]. In fact, as pointed out in [33], the hierarchical structure of language may be caused by our perception of reality, rather than the other way around. The case for a hierarchically organized world was made by Simon [42]:

> "On theoretical grounds we could expect complex systems to be hierarchies in a world in which complexity had to evolve from simplicity."

In addition, the evidence that nouns are easily changed into verbs (e.g., ship-shipped, bottle-bottled) [43] illustrates the same type of continuity in the signal space as well.

In any event, our solution is in line with the traditional Darwinian explanation to the evolution of the so-called irreducibly complex systems. Although the evolutionary game setting failed to evolve perfect compositional codes when the task was to produce a meaning-signal mapping by assimilating all meanings simultaneously, that setting proved successful when the meanings were created gradually.

V. Conclusion

Saussure's notion of language as a contract signed by members of a community to set arbitrarily the correspondence between words and meanings leads to unexpected obstacles to the evolution of efficient communication codes in the evolutionary language game framework. In fact, the fixation of a communication code in a population is a once-for-all decision—it cannot be changed even if a small fraction of the population acquires a different, more efficient code (see Fig. 7). The situation here is similar to the evolutionary stable strategies of game theory [27], the escape from which is only possible if all players change their strategies simultaneously. Since such concerted, global changes are not part of the rules of the language game, there seems to be no way for the population to escape from nonoptimal communication codes.

In fact, languages evolve. A branch of linguistics named glottochronology (the chronology of languages) suggests the rule of thumb that languages replace about 20% of their basic vocabulary every 1000 years [44]. The abovementioned difficulty of changing the communication code is not in the replacement of old signals by new ones, but in the assignment of different meanings to old signals and *vice versa*. Of course, this would not be an issue if the evolutionary language game could lead the population to the optimal code (a perfectly compositional code, in our case); our simulations have shown that it always gets stuck in one of the local maxima that plague the search space. To point out this difficulty was, in fact, the main goal of the present contribution.

Our view of compositionality as the evolutionary stage following the settlement of simpler, unstructured communication codes, and the search for a continuous path connecting these two stages, led us to the same type of difficulties researchers working on a similarly elusive problem—the origin of life—have been struggling with for more than three decades [39]. For example, although the coordinated work of distinct genes is germane to the emergence of cells, it is still not clear how such an assemblage could be formed and maintained starting from selfish genes (see [45] for a review). In that sense, by exposing the obstacles to explain compositionality from an evolutionary perspective, our work follows the same research vein that lead to the present understanding of prebiotic evolution.

The solution we put forward to this conundrum is a conservative one—we cannot explain the emergence of the entire meaning-signal mapping that displays the required compositional property via natural selection, but it is likely that the mapping was formed gradually with the addition of one meaning

at each time. This gradual procedure, that we term incremental meaning creation, leads indeed to fully compositional codes (see Fig. 8). It would be interesting to verify whether alternative, less conservative solutions such as the spatial localization of the agents, less than perfect metrics in meaning space, or the structuring of the population by age could lead to the dissolution of the language contract and so open an evolutionary pathway to more efficient communication codes.

REFERENCES

[1] F. de Saussure, *Course in Generol Linguistics*. New York: McGraw-Hill, 1966, Translated by Wade Baskin.

[2] M. A. Nowak, N. L. Komarova, and P. Niyogi, "Computational and evolutionary aspects of language," *Nature*, vol. 417, pp. 611–617, 2002.

[3] J. R. Hurford, "Biological evolution of the Saussurean sign as a component of the language acquisition device," *Linguo*, vol. 77, pp. 187–222, 1989.

[4] B. J. MacLennan, "Synthetic ethology: An approach to the study of communication," in *Artificiol Life II*, C. G. Langton, C. Taylor, J. Doyne Farmer, and S. Rasmussen, Eds. Reading, MA: Addison-Wesley, 1991, vol. X, SFI Studies in the Sciences of Complexity, pp. 631–658.

[5] K. Smith, S. Kirby, and H. Brighton, "Iterated learning: A framework for the emergence of language," *Artif. Life*, vol. 9, pp. 371–386, 2003.

[6] H. Brighton, "Compositional syntax from cultural transmission," *Artif. Life*, vol. 8, pp. 25–54, 2002.

[7] H. Brighton, K. Smith, and S. Kirby, "Language as an evolutionary system," *Phys. Life Rev.*, vol. 2, pp. 177–226, 2005.

[8] M. A. Nowak and D. C. Krakauer, "The evolution of language," *Proc. Nat. Acad. Sci. USA*, vol. 96, pp. 8028–8033, 1999.

[9] W. Zuidema, "Optimal communication in a noisy and heterogeneous environment," in *Proc. 7th Eur. Conf. Arti. Life (ECAL), Advances in Artificiol Life*, W. Banzhaf, T. Christaller, P. Dittrich, J. T. Kim, and J. Ziegler, Eds., 2003, vol. 2801, Lecture Notes in Artificial Intelligence, pp. 553–563.

[10] W. C. Allee, *Animal Aggregotions. A Study in General Sociology*. Chicago, IL: Univ. Chicago Press, 1931.

[11] F. Courchamp, T. Clutton-Brock, and B. Grenfell, "Inverse density dependence and the Allee effect," *Trends Ecol. Evol.*, vol. 14, pp. 405–410, 1999.

[12] D. Bickerton, *Language & Species*. Chicago, IL: Univ. Chicago Press, 1990.

[13] I. Ulbaek, "The origin of language and cognition," in *Approaches to the Evolution of Language*, J. R. Hurford, M. Studdert-Kennedy, and C. Knight, Eds. Cambridge, U.K.: Cambridge Univ. Press, 1998, pp. 30–43.

[14] L. Steels, "Perceptually grounded meaning creation," in *Proc. 2nd Int. Conf. Multi-Agent Syst.*, M. Tokoro, Ed., 1996, pp. 338–344.

[15] J. F. Fontanari and L. I. Perlovsky, "Meaning creation and modeling field theory," in *Proc. Int. Conf. Integr. Knowl. Intensive Multi-Agent Syst.*, C. Thompson and H. Hexmoor, Eds., 2005, pp. 405–410.

[16] M. A. Nowak, J. B. Plotkin, and D. C. Krakauer, "The evolutionary language game," *J. Theor. Biol.*, vol. 200, pp. 147–162, 1999.

[17] J. B. Plotkin and M. A. Nowak, "Language evolution and information theory," *J. Theor. Biol.*, vol. 205, pp. 147–159, 2000.

[18] J. F. Fontanari and L. I. Perlovsky, "Evolution of communication in a community of simple-minded agents," in *Proc. Int. Conf. Integr. Knowl. Intensive Multi-Agent Syst.*, C. Thompson and H. Hexmoor, Eds., 2005, pp. 285–290.

[19] D. K. Lewis, *Convention: A Philosophical Study*. Cambridge, MA: Harvard Univ. Press, 1969.

[20] R. Dawkins and J. R. Krebs, "Animal signals: Information or manipulation?," in *Behavioural Ecology: An Evolutionary Approach*, J. R. Krebs and N. B. Davies, Eds. Oxford, U.K.: Blackwell, 1978, pp. 282–309.

[21] J.-L. Dessalles, "Altruism, status and the origin of relevance," in *Approaches to the Evolution of Language*, J. R. Hurford, M. Studdert-Kennedy, and C. Knight, Eds. Cambridge, U.K.: Cambridge Univ. Press, 1998, pp. 130–147.

[22] R. L. Trivers, "The evolution of reciprocal altruism," *Quart. Rev. Biol.*, vol. 46, pp. 35–57, 1971.

[23] W. T. Fitch, "Kin selection and mother tongues: A neglected component in language evolution," in *Evolution A of Communication Systems: A Comparative Approach*, K. Oller and U. Griebel, Eds. Cambridge, MA: MIT Press, 2004, pp. 275–296.

[24] G. M. Burghardt, "Defining communication," in *Communication by Chemical Signals*, J. W. Johnston, Jr., J. D. G. Moulton, and A. Turk, Eds. New York: Appleton-Century-Crofts, 1970, pp. 5–18.

[25] M. D. Hauser, *The Evolution of Communication*. Cambridge: MIT Press, 1996.

[26] M. Oliphant, "The dilemma of Saussurean communication," *BioSystems*, vol. 37, pp. 31–38, 1996.

[27] J. M. Smith, *Evolution and the Theory of Games*. Cambridge, U.K.: Cambridge Univ. Press, 1982.

[28] M. Mitchell, *An Introduction to Genetic Algorithms*. Cambridge, MA: MIT Press, 1996.

[29] M. A. Nowak, J. B. Plotkin, and V. A. A. Jansen, "The evolution of syntactic communication," *Nature*, vol. 404, pp. 495–498, 2000.

[30] M. A. Nowak, N. L. Komarova, and P. Niyogi, "The evolution of universal grammar," *Science*, vol. 291, pp. 114–118, 2001.

[31] S. Pinker, *The Language Instinct*. Baltimore, MD: Penguin, 1994, p. 152.

[32] L. A. Petitto, "Language in the prelinguistic child," in *Language Acquisition: Core Readings*, P. Bloom, Ed. Cambridge, MA: MIT/Bradford Press, 1994.

[33] P. T. Schoenemann, "Syntax as an emergent characteristic of the evolution of semantic complexity," *Minds ond Mochines*, vol. 9, pp. 309–346, 1999.

[34] J. Aitchinson, *Words in the Mind: An Introduction to the Mental Lexicon*. Oxford, U.K.: Blackwell, 1994.

[35] W. J. Ewens, *Mathematical Populotion Genetics*, 2nd ed. New York: Springer, 2004.

[36] D. Fudenberg and J. Tirole, *Game Theory*. Cambridge, MA: MIT Press, 1991.

[37] R. M. Seyfarth, D. L. Cheney, and P. Marler, "Monkey responses to three different alarm calls: Evidence of predator classification and semantic classification," *Science*, vol. 210, pp. 801–803, 1980.

[38] M. Lachmann and C. T. Bergstrom, "The disadvantage of combinatorial communication," *Proc. R. Soc. Lond. B*, vol. 271, pp. 2337–2343, 2004.

[39] M. Eigen, "Self-organization of matter and the evolution of biological macro-molecules," *Naturwissenschaften*, vol. 58, pp. 465–523, 1971.

[40] H. A. Simon, "On a class of skew distribution functions," *Biometriko*, vol. 42, pp. 425–440, 1955.

[41] L. B. Levitin, B. Schapiro, and L. I. Perlovsky, "Zipf's law revisited: Evolutionary model of emergent multiresolution classification," in *Proc. Conf. Intell. Syst. Semiotics*, Gaithersburg, MD, 1996, vol. 1, pp. 65–70.

[42] H. A. Simon, *The Sciences of the Artificial*. Cambridge, MA: MIT Press, 1996.

[43] P. J. Hopper and S. A. Thompson, "The discourse basis for lexical categories in universal grammar," *Language*, vol. 60, pp. 703–752, 1984.

[44] C. Renfrew, *Archaeology ond Language*. London, U.K.: Pimlico, 1998.

[45] J. M. Smith and E. Szathmáry, *The Major Tronsitions in Evolution*. San Francisco, CA: Freeman, 1995.

José Fernando Fontanari received the Ph.D. degree in physics from the University of São Paulo (USP), São Paulo, Brazil, in 1988

He is a Professor of Theoretical Physics at USP. He worked as a Postdoctoral Researcher at the California Institute of Technology in J. Hopfield's group in 1989. The author of more than 80 papers in international journals with selective editorial policy has served as a Member of the Editorial Board of *Network: Computation in Neural Systems* from 1990–1992 and currently serves on the Editorial Board of *Physics of Life Reviews* and *Theory in Biosciences*. His research focuses on the application of concepts and analytical tools from physics to problems in biology, in particular, population dynamics and evolutionary theory.

Dr. Fontanari was elected Fellow of the Institute of Physics (U.K.) in August 2004.

Leonid I. Perlovsky (M'85–SM'95) is a Visiting Scholar at Harvard University, Cambridge, MA, and Principal Research Physicist and Technical Advisor at the Air Force Research Laboratory, Hanscom AFB, MA. He is Program Manager for DOD Semantic Web program and leads several research projects. From 1985 to 1999, he served as Chief Scientist at Nichols Research, a $0.5B high-tech DOD contractor leading the corporate research in intelligent systems, neural networks, sensor fusion, and target recognition. He served as Professor at Novosibirsk University and New York University, and participated as a principal in commercial startups developing tools for text understanding, biotechnology, and financial predictions. His company predicted the market crash following 9/11 a week before the event, detecting activities of Al Qaeda traders, and later helped SEC looking for these guys. He delivered invited keynote plenary talks and tutorial lectures worldwide, published more then 250 papers, 7 book chapters, and authored a monograph *Neural Networks and Intellect* (Oxford University Press, 2001) (currently in the 3rd printing). *The Knowledge Instinct* is being published in 2007 by Basic Books.

Dr. Perlovsky received the IEEE Distinguished Member of Boston Section Award and the International Neural Network Society Gabor Award. He serves as Associate Editor for the IEEE TRANSACTIONS ON NEURAL NETWORKS, Editor-at-Large for *Natural Computations*, and Editor-in-Chief for *Physics of Life Reviews*. He organizes conferences on Computational Intelligence, and Chairs the IEEE Boston Computational Intelligence Chapter.

Inverse density dependence in the evolution of communication

José F. Fontanari *

Instituto de Física de São Carlos, Universidade de São Paulo, Caixa Postal 369, 13560-970 São Carlos SP, Brazil

Leonid I. Perlovsky

Harvard University, 33 Oxford St, Rm 336, Cambridge MA 02138 and Air Force Research Laboratory, 80 Scott Drive, Hanscom Air Force Base, MA

Abstract

Structured meaning-signal mappings, i.e., mappings that preserve neighborhood relationships by associating similar signals with similar meanings, are advantageous in an environment where signals are corrupted by noise and sub-optimal meaning inferences are rewarded as well. The evolution of these mappings, however, cannot be explained within a traditional language evolutionary game scenario in which individuals meet randomly because the evolutionary dynamics is trapped in local maxima that do not reflect the structure of the meaning and signals spaces. Here we use a simple game theoretical model to show analytically that when individuals adopting the same communication code meet more frequently than individuals using different codes – a result of the spatial organization of the population – then advantageous linguistic innovations can spread and take over the population. In addition, we report results of simulations in which an individual can communicate only with its K nearest neighbors and show that the probability that the lineage of a mutant that uses a more efficient communication code becomes fixed decreases exponentially with increasing K. These findings support the mother tongue hypothesis that human language evolved as a communication system used among kin, especially between mothers and offspring.

Key words: Evolution of communication; Population dynamics; Inverse density dependence; Evolutionary games

1. Introduction

The notion that words compete and languages evolve in analogy to individuals and populations was already familiar in the nineteenth century as expressed in this quotation by the famous Darwin contemporary philologist Max Müller, "A struggle for life is constantly going on amongst the words and grammatical forms in each language. The better, the shorter, the easier forms are constantly gaining the upper hand, and they owe their success to their own inherent virtue" (Radick , 2002). A more suitable analog to language, however, is that of a parasitic species since language does not exist without speakers, just like parasites do not exist without hosts (Mufwene , 2001). In fact, the propagation of linguistic innovations through a population depends solely on the interaction between individuals and, as we will show here, the meeting practices of the speakers can hamper or facilitate the spread of new words or grammatical forms, regardless of their worth.

The debate on language evolution has centered mainly on the apparent gap between animal communication systems and human language (see, e.g., Pinker and Bloom (1990)). In fact, (non-human) animals use non-syntactic or holistic communication codes, i.e., signals refer to whole situations, in contrast to human language which is characterized by signals formed by discrete components that have their own meaning. As pointed out by Deacon (1997), no "simple" language which uses some elementary form of syntax or words combination has ever been found either in humans or in animals (see, however, Gordon (2004) for a possible exception - the puzzling language of the Pirahã people which lacks subordinate clauses as well as words associated with time, color and numbers). This discontinuity is behind the notion of a "language organ" exclusive of the human species which was originally designed to carry out combinatorial calculations (Chomsky , 1972; Fodor , 1983). According to Chomsky (1972), language is an example of true emergence - the appearance of a qualitatively different phe-

* Corresponding author. Fax:+55-16-33739877

Email addresses: fontanari@ifsc.usp.br (José F. Fontanari), Leonid.Perlovsky@hanscom.af.mil (Leonid I. Perlovsky).

nomenon at a specific stage of complexity of organization. The burden of identifying the selective pressures accountable for the emergence of syntax falls to those who hold the biological-oriented perspective that human language has evolved gradually from a simpler precursor - a proto-language - by means of the usual natural selection process. The demands of the social life of early hominids have been pointed as a probable source of selective pressures for the evolution of syntactic communication (Dunbar , 1996).

Even the evolution of simple holistic communication, which can be viewed as a one-to-one mapping between meanings and signals, has to confront some fundamental difficulties (Dawkins and Krebs , 1978; Fitch , 2004). In fact, from the perspective of the signaller, passing useful information to another individual is an altruistic act and so its maintenance in nature is problematic, whereas from the receiver viewpoint deciding whether a signal is honest (in the sense of conveying accurate information) or not is a difficult problem, the solution of which is thought to depend on the cost paid by the signaller to emit the signal (Zahavi , 1993). This is the essence of the "handicap principle", namely, honest signals are retained only when the signaller pays a high cost when emitting them (Zahavi , 1975). The relevance of this principle to the evolution of communication, however, has been defied by Noble (2000) who showed that a necessary condition for efficient communication to evolve is that both sender and receiver are benefited equally in the case of mutual understanding. By an efficient communication code we mean a Saussurean communication system that maps meanings unambiguously onto signals and then back into the original meanings (Hurford , 1989; Oliphant , 1996).

Rather than focusing on the evolution of Saussurean communication (see Hurford (1989); MacLennan (1991); Nowak and Krakauer (1999); Nowak et al (1999); Oliphant (1996); Noble (2000) for work on this line), in this paper we admit that one such a code is already established in the population and study the conditions under which a more robust communication system can take over. The breaking of the degeneracy between distinct Saussurean codes – essentially the $n!/(n-m)!$ different manners to associate m meanings to $n \geq m$ signals – is achieved by introducing errors in the perception of signals as well as by rewarding the inference of meanings close to the intended ones (Nowak and Krakauer , 1999; Zuidema , 2003; Fontanari and Perlovsky , 2007). This amounts to considering structured meaning-signal mappings in which neighborhood relationships are preserved (see Sect. 2).

We take up the evolutionary language game approach (Nowak et al , 1999) to study the competition between two communication codes or strategies: a perfectly structured meaning-signal mapping (strategy 1) and a random meaning-signal mapping (strategy 2). This study is primarily motivated by the failure of the traditional language game scenario to explain the evolution of structured communication codes starting from a population composed of individuals who use distinct communication codes, so the chance that a signal emitted by an individual is correctly interpreted by another individual is $1/m$ (Fontanari and Perlovsky , 2007). This is so because the evolutionary dynamics is very likely to get trapped in the local maxima – the random meaning-signal mappings – and once a communication code is fixed in the population it cannot be changed even if a small fraction of the population adopts the more efficient structured code. This is essentially the Allee effect (Allee , 1931) of population dynamics that asserts that intraspecific cooperation might lead to inverse density dependence, resulting in the extinction of some (social) animal species when their population size becomes small. Of course, this effect is germane to the outcome of biological invasions involving such species.

Instead of using the genetic algorithm to simulate the population dynamics, here we use an analytical approach based on the game theoretical formulation of Eshel and Cavalli-Sforza (1982), which allows us to derive explicit conditions for the minimum size of the population that adopts a structured code to invade an established population of individuals adopting a sub-optimal communication system. In particular, we show that useful linguistic innovations can spread and take over the population if the meeting of individuals using the same communication strategy is more likely than the encounter of individuals using different strategies – a natural consequence of imposing a spatial structure to the population since individuals are more likely to communicate with those close to them than with those farther away. Additional support to this finding is obtained through the explicit simulation of a spatially organized population in which the individuals can interact with their K-nearest neighbors only. Our findings support the "mother tongue" assumption that human language evolved as a communication system used among kin, especially between parents and their offspring (Fitch , 2004).

2. Meaning-signal mapping

Here we adopt the view that language is a mapping between meanings (or objects) and signals. In most previous studies of evolutionary language games this mapping is structureless or random: the metrics (if any) of the meaning and signal spaces play no role in the properties of the mapping and hence on the nature of the evolved communication codes (Hurford , 1989; MacLennan , 1991; Nowak and Krakauer , 1999; Nowak et al , 1999; Oliphant , 1996; Noble , 2000). This contrasts with a more recent approach that put emphasis on the properties of the meaning-signal mapping and, in particular, focus on structured mappings that preserve neighborhood relationships, i.e., nearby meanings in the meaning space are likely to be associated to nearby signals in signal space (Smith et al. , 2003; Zuidema , 2003; Brighton et al. , 2005; Fontanari and Perlovsky , 2007).

This notion of structured mappings seems contradictory to the well-established fact that the relation between a word (signal) and its meaning is arbitrary (Pettito , 1994). In

fact, as pointed out by Pinker (1994) "babies should not, and apparently do not, expect cattle to mean something similar to battle, or singing to be like stinging, or coats to resemble goats". On the other hand, a code that preserves neighborhood relationships is clearly advantageous in an environment where signals are likely to be altered by noise. Consider for instance the Vervet monkey alarm calls (Seyfarth et al. , 1980): misinterpreting a snake alarm for a leopard one, and hence running to a tree instead of standing up and looking in the grass, is clearly much better than misinterpreting it for an eagle call. In addition, sentences like *John walked* and *Mary walked* have parts of their semantic representation in common (someone performed the same act in the past) and so the meaning of these sentences must be close in the meaning space. Since both sentences contain the word *walked* they must necessarily be close in signal space as well (Smith et al. , 2003; Brighton et al. , 2005). It should be noted that the very notion of meaning similarity in contraposition to meaning identity is a highly controversial issue in cognitive science (see, e.g., Churchland (1998); Fodor and Lepore (1999); Abbott (2000). However, within a connectionist perspective in which meanings are neural activation patterns, the concept of meaning similarity follows naturally. In this contribution, we take the stand that in simple (nonhuman) communication structured meaning-signal mappings are likely to be relevant even at the elementary level of the object-word pairing, whereas in human language these mappings may play a role at the meaning-sentence level only.

We represent the signals (sentences or words) as well as the meanings by single symbols (labels) - only the "distance" between these entities will reflect the complex inner structure of the signal and meaning spaces. For instance, suppose there are only two words that we represent, without lack of generality, by 0 and 1 so that a binary sequence or, equivalently, its decimal representation stands for any sentence in this language. Here the relevant distance between two such sentences is the Hamming distance rather than, e.g., the result of the subtraction between their labeling integers. This notion, of course, generalizes trivially to the case where the sentences are composed of more than two types of words. As pointed out before, the representation of meanings is a much vaguer issue, but within a connectionist stand we can think of meanings also as patterns of 1s and 0s representing the arrangement of active and inactive neurons in the neural region activated by the signal.

For simplicity, in this paper we consider the case where both signals and meanings are represented by integer numbers and the relevant distance in both signal and meaning space is the result of the usual subtraction between integers. In addition, we consider the case where the number of signals equals the number of meanings $m = n$. Figure 1 illustrates a structured meaning-signal mapping in the case of $n = 5$ signals. For n signals there are only $2n$ structured mappings out of the $n!$ possible mappings. A random mapping is obtained simply by assigning meanings to signals randomly. A quantitative measure of the structure of a mapping is given by the degree to which the distances between all the possible pairs of meanings correlates with the distance between their corresponding pairs of signals, a quantity known as Pearson's correlation coefficient (Brighton et al. , 2005). Since here we will focus on the competition between two communication strategies given *a priori* – structured and random mapping – we will not consider these partially structured mappings which play a fundamental role when the issue is the emergence of structured mappings from an initial population of random mappings (Fontanari and Perlovsky , 2007).

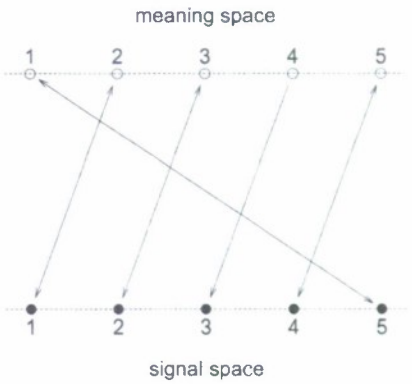

Fig. 1. Illustration of a meaning-signal mapping for $n = 5$. The integers here are viewed as labels for complex entities (e.g., sentences). The metric of the signal space is such that signal 3 is one unit distant from signals 2 and 4. This space has periodic boundary conditions so that signal 5 is one unit distant from signal 1. This metric applies to the meaning space as well. Because nearby meanings in the meaning space are associated to nearby signals in signal space this is a structured mapping.

A mapping that preserves the topology of the meaning and signal spaces was termed a compositional mapping in previous works (Smith et al. , 2003; Zuidema , 2003; Brighton et al. , 2005; Fontanari and Perlovsky , 2007). Here we use the term structured mapping instead, to avoid confusion with the well-established concept of compositionality which is defined as the property that the meaning of a complex expression is determined by the meanings of its parts and the rules used to combine them. In fact, Fodor and Lepore (1999) even claim that the notion of meaning similarity excludes the possibility of compositionality (see, however, Abbott (2000)). In an artificial scenario in which there is a prescription to derive the meaning of the whole given the meaning of the elementary parts, however, there is direct connection between structured and compositional meaning-signal mappings since in this case the distance between any two composite meanings could be inferred by comparing their components and, consequently, by introducing a metric in the meaning space.

3. Strategy payoffs

To explore the structure of the meaning-signal mapping (see Fig. 1) we must admit the possibility of errors in the perception of the signals as well as the alternative of rewarding the inference of meanings close but not equal to the meaning intended by the signaller.

It is reasonable to assume that in the earlier stages of the evolution of communication the signals were likely to be noisy and so they could be easily mistaken for each other. The relevance of the structure of the signal space becomes apparent when we note that the closer two signals are, the higher the chances that they are mistaken for each other. In particular, here we will consider the simple case in which there is a nonzero probability $\epsilon \in [0, 1/2]$ that a signal, say signal j, be mistaken for one of its nearest neighbors $j-1$ or $j+1$. So, in the example of Fig. 1, signal 5 can be mistaken for signal 4 with probability $\epsilon/2$ or for signal 1 with probability $\epsilon/2$. Of course, the probability that a signal is not corrupted by noise is $1-\epsilon$.

The individuals in the population can adopt either strategy 1 (structured meaning-signal mapping) or strategy 2 (random meaning-signal mapping). The interaction – a communication event – between a pair of individuals, say individuals I and J, comprises two stages: first I plays the role of signaller (so J is the receiver) and then I and J exchange roles. Both individuals receive the same payoff value. In particular, we assume that both signaller and receiver are rewarded with $1/2$ unity of payoff whenever the receiver interprets correctly the meaning of the emitted signal. In addition, both agents are rewarded with $r/2$ unity of payoff, where $r \in [0, 1]$, if the inferred meaning is one of the nearest neighbors of the intended meaning. We note that giving value to decisions which are not the best ones is a common assumption in decision and game theory (Fudenberg , 1991) and, it seems to be consistent with what is actually observed in nature since, as illustrated by the Vervet monkey alarm calls example, not every misinterpretation is equally harmful (see, e.g., Zuidema (2003)). The factors $1/2$ appear here because, as pointed out before, a communication event comprises two stages in which the individuals interchange the roles of signaller and receiver. So, both individuals gain 1 unity of payoff in case communication was successful in both stages.

Next we calculate the average payoff accrued to a pair of individuals during a communication event. First, let us consider the interaction between two individuals who both have strategy 1. The average payoff of the individual playing the signaller is $(1-\epsilon) \times 1/2 + \epsilon \times r/2$ which, by symmetry, happens to be the same average payoff it receives when playing the receiver role. Hence

$$F_{11} = 1 - \epsilon\,(1 - r)\,. \tag{1}$$

In the case both individuals have strategy 2, the average payoff of the signaller is $(1-\epsilon) \times 1/2 + \epsilon \times 2/(n-1) \times r/2$ where the factor $2/(n-1)$ accounts for the fact that the reward $r/2$ is obtained only if the inferred meaning is one of the two neighbors of the correct meaning. Hence the average payoff accrued to both individuals in a communication event is

$$F_{22} = 1 - \epsilon + \frac{2\epsilon}{n-1} r. \tag{2}$$

This reasoning is valid for $n > 2$ only: for $n = 2$ each meaning has a single neighbor and so the correct expression is $F_{22} = 1 - \epsilon\,(1 - r)$. Finally, in the case the individuals have different strategies the probability the receiver infers correctly the signaller intentions is simply $1/n$ and the probability that it infers a meaning which is a neighbor of the intended one is $2/(n-1)$. The average payoff of this communication event is then

$$F_{12} = \frac{1}{n} + \frac{2r}{n-1}. \tag{3}$$

and $F_{21} = F_{12}$.

For $n \geq 3$ we have $F_{11} \geq F_{22}$ where the equality holds for $n = 3$ as well as for the trivial cases $\epsilon = 0$ or $r = 0$. In addition, $F_{11} > F_{12}$ for $n > 2$. We will show in the following section that, except for $n = 4$ and ϵ close to its maximum value $1/2$, we have $F_{22} > F_{12}$. These inequalities are important to determine the local stability of the two strategies.

4. Population dynamics

As pointed by Ferdinand de Saussure "language is not complete in any speaker; it exists only within a collectivity... only by virtue of a sort of contract signed by members of a community" (Saussure , 1966). Translated into the biological jargon, this assertion means that language is not the property of an individual, but the extended phenotype of a population (Nowak et al , 2002). So a suitable approach to language evolution must take into account the population dynamics. In what follows we build on the game theoretical formulation of Eshel and Cavalli-Sforza (1982) to investigate analytically the evolution of structured communication codes.

Let $x \in [0, 1]$ be the proportion of individuals in a population of infinite size that use the structured communication code (strategy 1). To calculate the expected payoff of individuals adopting a particular strategy we need to make some assumption about the frequency of encounters between any two individuals. Let u_{ij} with $i, j = 1, 2$ be the probability that an individual using strategy i encounters an individual that uses strategy j. Since the game rules are such that an individual must encounter a partner to interact with, we have $u_{i1} + u_{i2} = 1$ for $i = 1, 2$. In addition, since the average number of encounters between individuals using different strategies can be written either as $x u_{12}$ or $(1-x)\,u_{21}$ we have the equality $u_{12}/u_{21} = (1-x)/x$. Hence a single encounter probability, say u_{11}, determines all other encounter probabilities: $u_{12} = 1 - u_{11}$, $u_{21} = x\,(1 - u_{11})\,/\,(1-x)$, and $u_{22} = (1 - 2x + x u_{11})\,/\,(1-x)$.

In the case encounters are random and independent of the communication code we have $u_{11} = x$ so that $u_{12} = u_{22} = 1 - x$ and $u_{21} = x$.

The expected payoff for individuals using strategy $i = 1, 2$ is $F_i(x) = u_{i1}F_{i1} + u_{i2}F_{i2}$ or, explicitly,

$$F_1(x) = F_{12} + (F_{11} - F_{12})\, u_{11}(x) \tag{4}$$

$$F_2(x) = F_{22} + (F_{12} - F_{22})\, \frac{x}{1-x}\, [1 - u_{11}(x)]\,. \tag{5}$$

A simple deterministic population dynamics model that describes the competition of the two strategies is obtained by assuming that the proportion of individuals using strategy 1 in generation $t+1$ is proportional to the relative payoff of that strategy in generation t, i.e.,

$$x_{t+1} = \frac{x_t F_1(x_t)}{x_t F_1(x_t) + (1 - x_t) F_2(x_t)} \equiv f(x_t)\,, \tag{6}$$

which essentially implies that mastery of a public communication system adds to the reproductive potential of the individuals (Hurford , 1989). This model is equivalent to the standard genetic algorithm (Mitchell , 1996) procedure with an infinite population size. As expected $x = 0$ and $x = 1$ are always fixed points of the recursion equation (6). The issue is to determine their stability and, in the case that both fixed points are stable, their basins of attraction. As usual, the condition for the stability of a fixed point x^* is simply $f'(x^*) < 1$ (see, e.g., Maynard Smith (1982)).

4.1. *Random encounters*

This is the typical scenario used in most computational models for the evolution of communication (Hurford , 1989; MacLennan , 1991; Nowak and Krakauer , 1999) and, in particular, Fontanari and Perlovsky (2007) have considered an agent-based simulation aiming at exploring the plausibility of the emergence of structured codes in a random encounter situation. As already mentioned, random encounters are described by $u_{11} = x$. The stability condition of the fixed point $x = 0$ associated to strategy 2 (random meaning-signal mapping), namely, $f'(0) < 1$ yields $F_{22} > F_{12}$ or, more explicitly,

$$r < \frac{n-1}{2}\left[1 - \frac{1}{n(1-\epsilon)}\right]. \tag{7}$$

Since $r \leq 1$ and $\epsilon \leq 1/2$ this condition is violated only if $n = 4$ and $\epsilon > 1/4$. Similarly, the fixed point $x = 1$, associated to strategy 1 (structured meaning-signal mapping), is stable provided $f'(1) < 1$, that leads to the condition $F_{11} > F_{12}$ which is satisfied for $n > 2$ regardless of the values of r and ϵ. In most cases (e.g., $n > 4$) the fixed points $x = 0$ and $x = 1$ are stable and so there is an inner unstable fixed point x_u that delimits the basins of attractions of the two stable fixed points. It is given by the condition $F_1(x_u) = F_2(x_u)$ which yields

$$x_u = \left(1 + \frac{F_{11} - F_{12}}{F_{22} - F_{12}}\right)^{-1}. \tag{8}$$

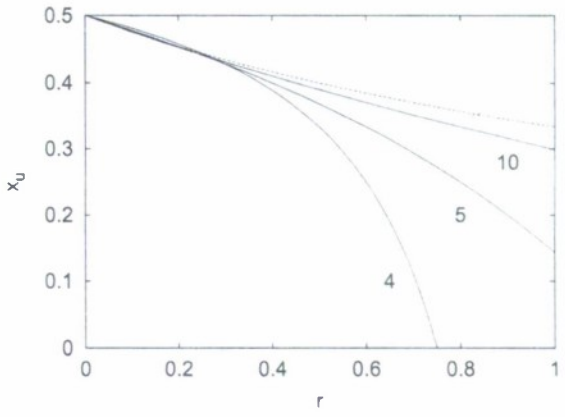

Fig. 2. Minimum fraction of individuals using structured codes necessary for this strategy to dominate the population in the case of random encounters for $\epsilon = 0.5$ and n as indicated in the figure. The dashed curve is the result for $n \to \infty$.

This quantity is the minimum initial fraction of individuals using strategy 1 above which this strategy dominates the population. Expression (8) corrects the estimate given in Fontanari and Perlovsky (2007). In Fig. 2 we illustrate the dependence of x_u on the parameters of the model. As already pointed out, since for $n = 4$ the Saussurean fixed point is unstable in the range $r > \frac{3}{2}[1 - 1/4(1-\epsilon)]$ we have $x_u = 0$ in this regime. For $n \to \infty$ we find $x_u = [2 + \epsilon r/(1-\epsilon)]^{-1}$.

4.2. *Nonrandom encounters*

Nonrandomness of encounters are usually modeled by imposing some spatial structure to the population in which the individuals are fixed to lattice sites and so can interact only with their nearest neighbors or then isolated in groups (see, e.g., Oliphant (1996); Noble (2000); Cangelosi (2001) for this type of approach within the evolution of communication context). An alternative formulation of nonrandom encounters which keeps the mathematics simple is to assume that the frequency of meetings between individuals using strategy 1 is

$$P_{11} = (1-m)\, x^2 + mx \tag{9}$$

where $m \in [0, 1]$ is the aggregation parameter (Wright , 1921; Eshel and Cavalli-Sforza , 1982). In fact, the probability that an individual using strategy 1 encounters another of its kind is $u_{11} = P_{11}/x = m + (1-m)\, x$, from which we obtain $u_{22} = m + (1-m)(1-x)$. Hence m represents the portion of the population that meets an individual of the same strategy, whereas the fraction $1 - m$ meets randomly. The situation of random encounters is obtained by setting $m = 0$.

Now the conditions for the stability of the fixed points $x = 0$ and $x = 1$ become $F_{22} > mF_{11} + (1-m)\, F_{12}$ and $F_{11} > mF_{22} + (1-m)\, F_{12}$, respectively. Since $F_{11} > F_{22}$ and $F_{11} > F_{12}$ for $n > 2$ the fixed point $x = 1$ is always sta-

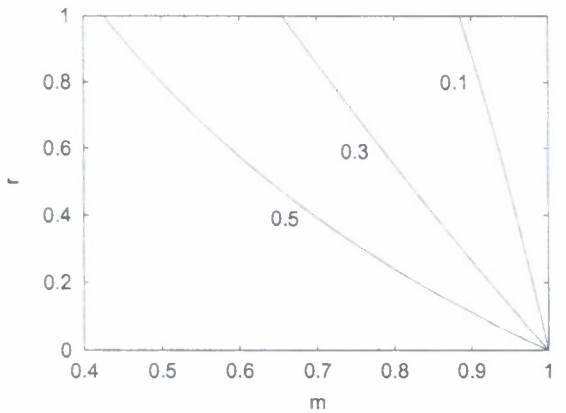

Fig. 3. Phase diagram in the plane (m, r) showing the regions where the fixed point $x = 0$ associated to the random meaning-signal mapping is unstable (above the curves) so the individuals using the structured communication code can dominate the population even when their initial frequency is vanishingly small. The parameters are $n = 10$ and $\epsilon = 0.1,\ 0.3,\ 0.5$ as indicated in the figure.

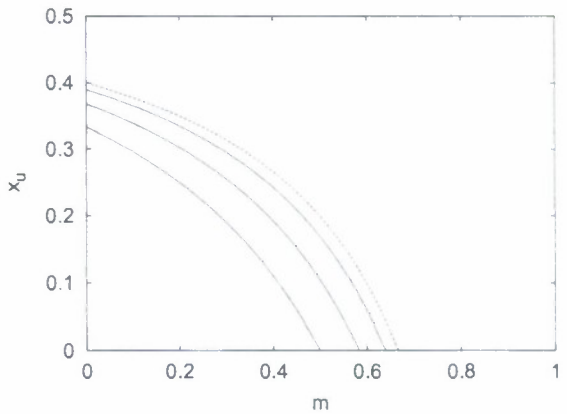

Fig. 4. Minimum fraction of individuals using structured communication necessary for this strategy to dominate the population in the case of nonrandom encounters for $\epsilon = r = 0.5$ and (solid curves from bottom to top) $n = 4$, 5, and 10. The dashed curve is the result for $n \to \infty$.

ble regardless of m. The situation for the fixed point $x = 0$, however, changes considerably, as illustrated in Fig. 3 that shows the regions of stability of this fixed point in the plane (m, r). Large values of m can, as expected, destabilize this fixed point. By setting the parameters so as to maximize the advantage of strategy 1, i.e., $r = 1$ and $\epsilon = 1/2$, we find that the stability of $x = 0$ is guaranteed provided that $m < m_s$ with $m_s = 1 - \frac{1}{2}n\,(n-3)\,/\left(n^2 - 4n + 1\right)$ for $n > 4$. Note that $m_s \in [1/6, 1/2]$ as n increases from 5 to ∞.

In the case both fixed points $x = 0$ and $x = 1$ are stable, the inner unstable fixed point is still given by the condition $F_1(x_u) = F_2(x_u)$ which now yields

$$x_u = \frac{1}{1-m}\left(1 - m\frac{F_{11} - F_{12}}{F_{22} - F_{12}}\right)\left(1 + \frac{F_{11} - F_{12}}{F_{22} - F_{12}}\right)^{-1}. \quad (10)$$

Figure 4, which exhibits the dependence of the threshold frequency x_u on the aggregation parameter m, reinforces the fact that $x_u = 0$ in the regions of the space of parameters where the fixed point associated with the random mapping strategy is unstable.

4.3. *Spatially structured populations*

In support to the findings of the previous section, here we report results of agent-based simulations where the spatial organization of the population is explicitly taken into account. In particular, we assume that N individuals are placed in equidistant sites on a ring (one individual per site), and each individual can interact with its Kth nearest neighbors only. The fully connected situation (i.e., an individual interacts with the $N - 1$ remaining individuals in the population) is recovered by setting $K = (N-1)/2$. As before, the individuals can use either strategy 1 or strategy 2 and the payoff resulting from their interactions are given by Eqs. (1)-(3). We recall that each interaction comprises two events in which the individuals exchange roles as signaler and receiver.

The fitness of an individual is evaluated by computing the total payoff it obtains when interacting with its K nearest neighbors. Once the fitness of all individuals are known, we compute the total fitness of the population and then the relative fitness of each individual. The next step is to choose a single individual, say I, for replication with probability proportional to its relative fitness. The copy of I then replaces one of the $2K + 1$ individuals that belong to the neighborhood of influence of I and I itself. The choice of the individual to be discarded is done randomly without regard to their fitness values. This selection procedure is essentially Moran model of population genetics (Ewens , 2004), except for the fact that the offspring is placed in the region of influence of its parent, resulting in a situation where neighbors are likely to be genetically similar (Oliphant , 1996). The repetition of this procedure for N times defines the time unit (one generation) of the dynamics. At the initial generation ($t = 0$), all individuals adopt strategy 2, except for a single mutant that uses strategy 1.

Figure 5 illustrates the time evolution of the fraction of individuals that use strategy 1 for four independent runs in the case an individual can interact only with its two first nearest neighbors ($K = 1$). We should note that these are not typical runs, since in a typical run the mutant lineage goes extinct in the very first generations. This figure highlights the stochastic character of the dynamics - the same initial setting can lead to very different outcomes, namely, the fixation or the extinction of the mutant lineage. To make this observation quantitative we record the outcome of 10^7 independent runs and present in Fig. 6 the fraction of them (P_s) that resulted in the fixation of the mutant lineage, i.e., of the structured communication code.

The most relevant information revealed by Fig. 6 is that the probability of invasion decreases exponentially with increasing K. In particular, for the data exhibited in the fig-

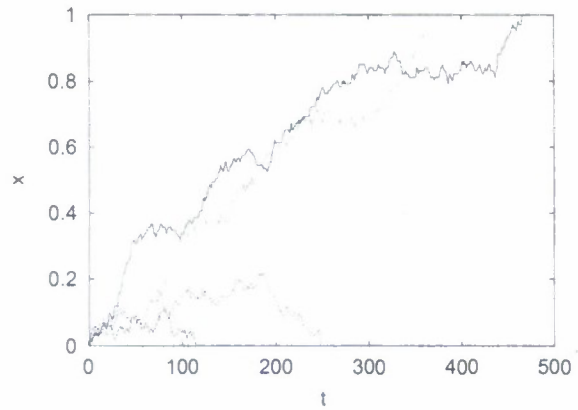

Fig. 5. Fraction of individuals that use strategy 1 in four population samples of $N = 101$ individuals placed in equidistant sites on a ring. Each individual can interact only with its first nearest neighbors ($K = 1$). The initial condition is $x_0 = 1/N$ and the parameters are $r = 1$, $\epsilon = 0.5$, and $n = 10$.

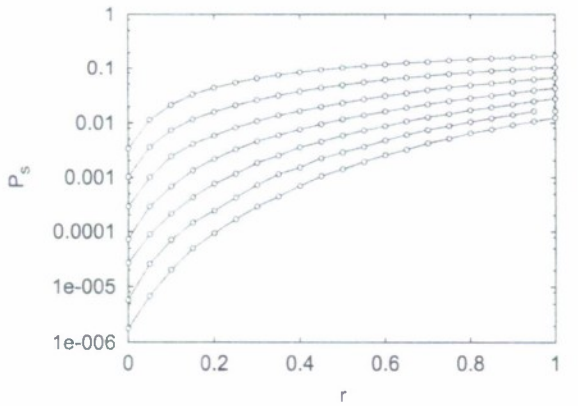

Fig. 6. Probability that the lineage of a single mutant that uses stategy 1 overtakes the resident population in a chain with $N = 101$ individuals, where each individual interacts with its $2K$ nearest neighbors. The parameters are $\epsilon = 0.5$, $n = 10$ and (top to bottom) $K = 1, 2, \ldots, 7$. The lines are guides to the eyes.

ure we find $P_s = a \exp(-bK)$ where $a \approx 0.03 + 0.47r - 0.23r^2$ is an increasing function of $r \in [0, 1]$ whereas $b \approx 1.2(1 - r) + 0.46r^2$ decreases with increasing $r \in [0, 1]$. The results for different values of the noise parameter ϵ exhibit the same qualitative behavior. In addition, in the range of K considered here, we have found that the fixation probability P_s is practically insensitive to the population size N. Hence in agreement with the findings of the previous section, the aggregation of individuals using the same communication system is ultimately the mechanism that lead to the spread of advantageous linguistic innovations in a population.

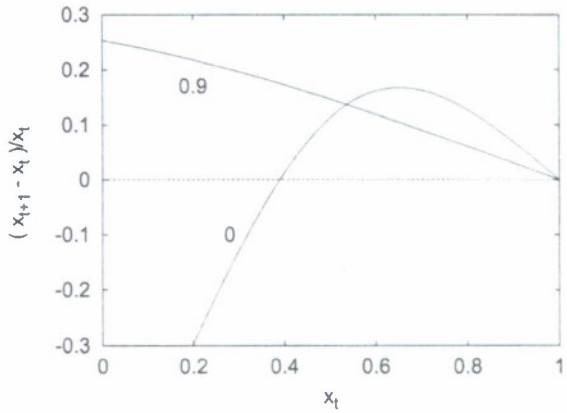

Fig. 7. Per capita growth rate of individuals using the structured communication code as function of the fraction of the population which adopts that strategy for $m = 0$ and $m = 0.9$ as indicated in the figure. The parameters are $n = 10$ and $\epsilon = r = 0.5$.

5. Discussion

Understanding how innovations that increase the expressive power of individuals can spread through a population and eventually become fixed is the essence of any evolutionary explanation to language evolution. However, the finding that the adoption of any particular trait (a structured communication code, in the present context) is better for a population, in the sense it yields an higher overall payoff, is no guarantee that such trait will actually spread in the population. As pointed out by Cavalli-Sforza and Feldman (1983), since communication takes place between two or more individuals, the selective process is frequency dependent and so communication cannot evolve in a simple scenario in which the individuals meet randomly. Those authors have argued that such obstacle can be removed, however, if the communication events occur predominantly within the family or among close relatives. Interestingly, this same idea reappeared about ten years later as the "mother tongue" scenario, which purports that language evolved as a communication system used among kin, especially between mothers and their offspring, so as to resolve the difficulties inherent to the altruistic behavior of the signaller when passing relevant information to the receiver (Fitch , 2004).

The paradox of the evolution of communication in a panmitic population is, in fact, an older idea: the very notion that intraspecific cooperation might lead to an inverse density dependence on the growth rate of some social animals is the essence of the Allee effect (Allee , 1931) (see Courchamp et al. (1999) for a review). Figure 7 illustrates the Allee effect, i.e., the inverse density dependence of the per capita growth rate in the case of random encounters ($m = 0$) and the usual density dependence in the case of strong assortative meetings ($m = 0.9$). Extinction is certain whenever the population of individuals who have strategy 1 reaches a frequency value for which the growth rate is negative.

In this contribution, we have expanded the work of

Cavalli-Sforza and Feldman (1983) by showing that the emergence of different communication codes, even when clearly advantageous in comparison with the code adopted by the resident population, is likely to be established only if some aggregation (or segregation) mechanism is acting on the population. There is vast evidence of this process in the linguistic literature, the more recent is probably the development of the Black English Vernacular dialect in black ghettos in America (Pinker , 1994).

Acknowledgments

This work was supported in part by the Air Force Office of Scientific Research, Air Force Material Command, USAF, under grant number FA9550-06-1-0202, and in part by CNPq and FAPESP, Project No. 04/06156-3. The U. S. Government is authorized to reproduce and distribute reprints for Governmental purpose notwithstanding any copyright notation thereon.

References

Abbott, B., 2000. Fodor and Lepore on Meaning Similarity and Compositionality. The Journal of Philosophy 97, 454-455.

Allee, W. C., 1931. Animal Aggregations. A Study in General Sociology. University of Chicago Press, Chicago.

Brighton, H., Smith, K., Kirby, S., 2005. Language as an evolutionary system. Phys. Life Rev. 2, 177-226.

Cangelosi, A., 2001. Evolution of Communication and Language Using Signals, Symbols, and Words. IEEE Trans. Evol. Comput. 5, 93-101.

Cavalli-Sforza, L. L., Feldman, M. W., 1983. Paradox of the evolution of communication and of social interactivity. Proc. Natl. Acad. Sci. USA 80, 2017-2021.

Chomsky, N., 1972. Language and mind. Harcourt Brace Jovanovich, New York.

Churchland, P.M., 1998. Conceptual Similarity across Sensory and Neural Diversity: The Fodor/Lepore Challenge Answered. The Journal of Philosophy 95, 5-32.

Courchamp, F., Clutton-Brock, T., Grenfell, B., 1999. Inverse density dependence and the Allee effect. Trends Ecol. Evol. 14, 405-410.

Dawkins, R., Krebs, J. R., 1978. Animal signals: information or manipulation? in: Krebs, J. R., Davies, N. B. (Eds.), Behavioural ecology: an evolutionary approach. Blackwel Scientific Publications, Oxford, UK, pp. 282-309.

Deacon, T. W., 1997. The Symbolic Species. W.W. Norton & Company, New York.

Dunbar, R., 1996. Grooming, Gossip, and the Evolution of Language. Harvard University Press, Cambridge, MA.

Eshel, I., Cavalli-Sforza, L. L., 1982. Assortement of encounters and evolution of cooperativeness. Proc. Natl. Acad. Sci. USA 79, 1331-1335.

Ewens, W. J., 2004. Mathematical Population Genetics, 2nd edition. Springer, New York.

Fitch, W. T., 2004. Kin selection and mother tongues: A neglected component in language evolution, in: Oller, K., Griebel U. (Eds.), Evolution of Communication Systems: A Comparative Approach, MIT Press, Cambridge, MA, pp. 275-296.

Fodor, J., 1983. The Modularity of Mind. MIT Press, Cambridge, MA.

Fodor, J., Lepore, E., 1999. All at Sea in Semantic Space: Churchland on Meaning Similarity. The Journal of Philosophy 96, 381-403.

Fontanari, J.F., Perlovsky, L.I., 2007. Evolving compositionality in evolutionary language games. IEEE Trans. Evol. Comput., doi:10.1109/TEVC.2007.892763

Fudenberg, D., Tirole, J., 1991. Game Theory. MIT Press, Cambridge, MA.

Gordon, P., 2004. Numerical cognition without words: Evidence from Amazonia. Science 306, 496-499.

Hauser, M. D., 1996. The Evolution of Communication. MIT Press, Cambridge, MA.

Hurford, J.R., 1989. Biological evolution of the Saussurean sign as a component of the language acquisition device. Lingua 77, 187-222.

MacLennan, B.J,, 1991. Synthetic ethology: an approach to the study of communication, in: Langton, C.G., Taylor, C., Doyne Farmer, J., Rasmussen, S. (Eds.), Artificial Life II, SFI Studies in the Sciences of Complexity, vol. X. Addison-Wesley, Redwood City, pp. 631-658.

Maynard Smith, J., 1982. Evolution and the Theory of Games. Cambridge University Press, Cambridge, UK.

Mitchell, M., 1996. An Introduction to Genetic Algorithms. MIT Press, Cambridge, MA.

Mufwene, S. S., 2001. The Ecology of Language Evolution. Cambridge University Press, Cambridge, UK.

Noble, J., 2000. Cooperation, competition and the evolution of prelinguistic communication, in: Knight, C., Studdert-Kennedy, M., Hurford, J. (Eds.), The Evolutionary Emergence of Language, Cambridge University Press, Cambridge, UK, pp. 40-61.

Nowak, M.A., Krakauer, D.C., 1999. The evolution of language. Proc. Natl. Acad. Sci. USA 96, 8028-8033.

Nowak, M. A., Plotkin, J. B., Krakauer, D. C., 1999. The Evolutionary Language Game. J. theor. Biol. 200, 147-162.

Nowak, M. A., Komarova, N. L., Niyogi, P., 2002. Computational and evolutionary aspects of language. Nature 417, 611-617.

Oliphant, M., 1996. The dilemma of Saussurean communication. BioSystems 37, 31-38.

Petitto, L.A., 1994. Language in the prelinguistic child, in: Bloom, P. (Ed.), Language acquisition: Core readings. MIT/Bradford Press, Cambridge.

Pinker, S., 1994. The Language Instinct. Penguin Press, London.

Pinker, S., Bloom, P., 1990. Natural language and natural selection. Behav. Brain Sci. 13, 707-784.

Radick, G., 2002. Darwin on Language and Selection. Selection 3, 7-16.

de Saussure, F., 1966. Course in General Linguistics. Translated by Wade Baskin. McGraw-Hill Book Company, New York.

Seyfarth, R.M., Cheney, D.L., Marler, P., 1980. Monkey responses to three different alarm calls: Evidence of predator classification and semantic classification. Science 210, 801-803.

Smith, K., Kirby, S., Brighton, H., 2003. Iterated Learning: a framework for the emergence of language. Artificial Life 9, 371-386,.

Wright, S., 1921. Systems of mating. III. Assortative mating based on somatic resemblance. Genetics 6, 144-161.

Zahavi, A., 1975. Mate selection: A selection for a handicap. J. Theor. Biol. 53, 205-214.

Zahavi, A., 1993. The fallacy of conventional signalling. Proc. Roy. Soc. London B 340, 227-230.

Zuidema, W., 2003. Optimal communication in a noisy and heterogeneous environment. Lecture Notes in Artificial Intelligence 2801, 553-563.

How communication can improve differentiation in the Modeling Field Theory framework

José F. Fontanari
Universidade de São Paulo, São Carlos, Brazil, fontanari@ifsc.usp.br
Leonid I. Perlovsky
Harvard University, Cambridge MA and The Air Force Research Laboratory, SN, Hanscom, MA
Leonid.Perlovsky@hanscom.af.mil

Abstract *— We propose a discrimination task scenario to study language acquisition in which an agent receives linguistic input from an external teacher, in addition to the sensory stimuli from the objects that make up the environment. The agent is endowed with the Modeling Field Theory (MFT) categorization mechanism, which enables it to identify a few objects (or categories) composed of hundreds of random pixels (instances). We show that the agent with language is capable of differentiating objects or categories that it could not distinguish without language.*

1. INTRODUCTION

The nature of the selective pressures accountable for the emergence of language has been object of passionate debates since the viewpoint that language evolved through natural selection has become dominant in the scientific community [1]. In this contribution we examine the suggestion that the selective pressures for language have come from the brute exigencies of survival, e.g., hunting, food gathering and predator detection (see [2], [3]). We refer the reader to Ref. [4] for the alternative and perhaps more popular stance that the leading role in language evolution was played instead by the demands of the social life of early hominids. Rather than focusing on the "origin" issue, here we take a more pragmatic view and consider these elementary survival needs as problems to be solved by the individuals (agents, in our case), and ask whether and how communication can improve their performances to solve a specific problem relevant to the individuals' endurance.

The specific task we consider in this contribution is the differentiation problem, i.e., how agents develop a more detailed knowledge of their surroundings. In fact, one possible advantage of communication is that a group of individuals with this capacity can perceive their environment beyond the limits of their senses: an individual unable to communicate can access its environment based on the information provided by its own senses only [5], [6]. Here we use the term differentiation as synonymous to discrimination. To be able to discriminate is to be able to judge whether two inputs are the same or different [7]. The ability to discriminate inputs depends on the constitution of iconic representations: same/different judgments are based on the sameness or difference of these representations, according to some inherent similarity measure. Discrimination is clearly independent of identification as one can discriminate things without knowing what they are [7]. For identification, icons must be reduced to a small set of invariant features that will reliably distinguish members of different categories. Recently we have shown how a novel adaptive approach to concept formation - Modeling Field Theory (MFT) [8] - can successfully integrate and implement these two tasks into a simple autonomous neural-networks-like scheme [9], [10]. Here we advance further and allow the agents endowed with a categorization system based on MFT to create symbolic or linguistic representations for members of a category, i.e., to name the category. This is a modest first step towards the ambitious program of fully integrating language and cognition [11], [12].

In the next section, we describe the environment in which the agent is embodied and embedded – the Umwelt in the ethologists' jargon - as well as the task posed to it. In section 3 we briefly review the MFT formalism within the context of the specific categorization problem addressed in this paper. In section 4 we present the results of the simulation of the MFT dynamic in the case the agent does not receive a linguistic input, and in section 5 we address the more interesting case where an external teacher names the objects as they are perceived by the agent. Finally, section 6 summarizes the main conclusions.

2. THE DIFFERENTIATION TASK

We assume that the world contains a certain number of objects whose features (e.g., color, smell, coordinates in a grid, texture, etc.) are modeled by overlapping sets of points drawn from Gaussian distributions. Figure 1 displays the particular instance we will consider in this paper. There are at least two (equivalent) interpretations

1-4244-0945-4/07/$25.00 ©2007 IEEE

for the problem we are about to tackle. First, we can view each point in the figure as representing two particular features of 600 objects which belong to six distinct categories. The issue is then to identify these categories. Note that identification presumes prior discrimination. Second, the 600 points displayed in Fig. 1 represent pixels of the image of six solid objects projected into a two-dimensional retina. The issue is then to identify the objects, a classic pattern recognition problem [13]. MFT has been applied to this type of problem with success for many years [8] (see [10] for use of MFT together with the Akaike Information Criterion [14] to estimate the number of objects in a scene). Here we adhere to this object-oriented interpretation. There are, however, a few issues we should mull over before embarking on the mathematical formulation of the agent-based model.

When talking about autonomous differentiation of objects we are implicitly assuming that the system knows somehow what an object is. This is a very difficult question as Marr's skeptical remark readily reminds us: "... all these things can be an object if you want to think of them that way, or they can be part of a larger object" [15]. The notion of object, however, is central to the understanding of how children acquire language. In that case, the problem seems to be solved by inborn mechanisms that implement the so-called principle of cohesion: an object is a connected and bounded region of matter that maintains its connectedness and boundaries when it is in motion [16]. It would be interesting to apply MFT to the categorization of sets of pixels in movement, since tracking of moving objects is one of the traditional applications of that method [16]. The bottom of the problem is that the notion of object must be explicitly built into the categorization scheme. In the MFT framework, this is done when we define *a priori* the range of concept models that are used to categorize the input data.

The environment is set so that an agent cannot categorize and identify all objects, because of the considerable overlap between them (see Figure 1). Inspired by the "mushroom" world scenario [5], [6], we allow the agent to receive from the environment an additional sensory input: a heard linguistic signal. Here, we assume that this signal is produced by an external teacher who has perfect knowledge of the agent's environment. In the following we will show that when the agent receives the linguistic input associated to the different objects, it can create new concept-models and so identify unambiguously all objects.

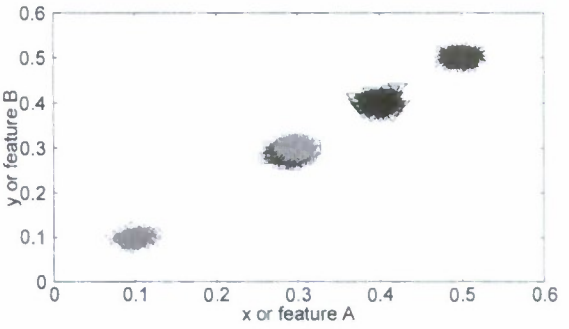

Figure 1: The six sets of 100 pixels represent the coordinates x and y of six objects or, alternatively, features A and B of six categories. The coordinates of each pixel are drawn from Gaussian distributions of means 0.1, 0.2, 0.29, 0.3, 0.4, and 0.5 and standard deviation 0.01.

3. MODELING FIELD THEORY

The basic idea behind Modeling Field Theory is the association between lower-level signals (e.g., inputs) and higher-level concept-models (internal representations) avoiding the combinatorial complexity inherent to such a task. This is achieved by using measures of similarity between concept-models and input signals together with a new type of logic, so-called fuzzy dynamic logic. We refer the reader to Perlovsky's book [8] for a complete presentation of MFT; here we particularize the general framework to the problem of categorizing the $N = 600$ pixels depicted in Figure 1. Each pixel is described by the pair of real variables (O_{1i}, O_{2i}) with $i = 1, \ldots, N$. Let us assume that there are M concept-models described by the pairs (S_{1k}, S_{2k}) with $k = 1, \cdots, M$ that should represent the original pixels. We define arbitrarily the following partial similarity measure between object i and concept k

$$l(i \mid k) = \prod_{e=1}^{2} \left(2\pi\sigma_{ek}^2\right)^{-1/2} \exp\left[-(O_{ei} - S_{ek})^2 / 2\sigma_{ek}^2\right] \qquad (1)$$

where, at this stage, the fuzziness σ_{ek} are parameters given *a priori*. We refer the reader to Ref. [17] for another application of MFT in the case of multi-component inputs. The goal is to find an assignment between models and objects such that the global similarity

$$L = \prod_i \sum_k l(i \mid k) \qquad (2)$$

is maximized. A fundamental role is played by the fuzzy association variables $f(k \mid i)$ defined by

$$f(k \mid i) = l(i \mid k) \Big/ \sum_{k'} l(i \mid k') \qquad (3)$$

which give a measure of the correspondence between object i and concept k relative to all other concepts k'. The maximization of the global similarity L can be achieved using the MFT mechanism of concept formation which is based on the following dynamics for the modeling fields

$$dS_{ek}/dt = \sum_i f(k|i)[\partial \log l(i|k)/\partial S_{ek}] \qquad (4)$$

for $e = 1,2$ and $k = 1,\cdots,M$. We note that although the coordinates x and y of a pixel are independent random variables, the two components of a modeling field are coupled dynamic variables. Actually, the term $f(k|i)$ in Eq. (4) couples not only S_{1k} and S_{2k} but also components of different modeling fields [17].

Proper adjustment of the fuzziness σ_{ek} during the evolution of the modeling fields allows the dynamics to converge to the maximum of the global similarity L. In particular, we decrease σ_{ek} according to the following prescription

$$\sigma_{ek}^2(t) = \sigma_{ae}^2 \exp(-\alpha_e t) + \sigma_{be}^2 \qquad (5)$$

with $\alpha_e = 5\times10^{-4}$, $\sigma_{be} = 0.03$, and $\sigma_{ae} = 0.5$ for $e = 1,2$. Note that these parameters are the same for all models $k = 1,\ldots,M$ and components $e = 1,2$. However, we will need different parameter values to stabilize the linguistic component $e = 3$, which we will introduce in the Section 5.

In what follows we will set $M = 6$ so Eq. (4) stands for a set of twelve nonlinear coupled equations, which are solved with Euler's method using the step-size $h = 10^{-5}$. As mentioned before, use of the MFT approach in conjunction with the Akaike Information Criteron has allowed us to design a categorization system that infers correctly the true number of objects in an environment similar to that exhibited in Figure 1 [10], but for the purposes of this paper, any choice of $M \geq 6$ will be satisfactory.

4. AGENT WITHOUT LANGUAGE

The problem is motivated by the inability of an agent to distinguish between the six objects that make up its environment. The difficulty, of course, is to distinguish between the two set of pixels centered at the coordinates (0.29,0.29) and (0.3,0.3) as shown in Figure 1. Figures 2 and 3 illustrates the time dependence of the two components, $e = 1$ (Feature A) and $e = 2$ (Feature B) of the modeling fields S_{ek}. Since Feature B is essentially equivalent to Feature A the associated modeling field components exhibit a very similar behavior pattern. What distinguish these components are simply their initial values, which were chosen randomly from the uniform distribution. The point of Figures 2 and 3 is to stress the rather expected failure of most categorization methods to distinguish highly overlapping objects. The agent is able to identify four of the six objects displayed in Figure 1 and, in addition, it can clearly discriminate the two overlapping objects from the other four.

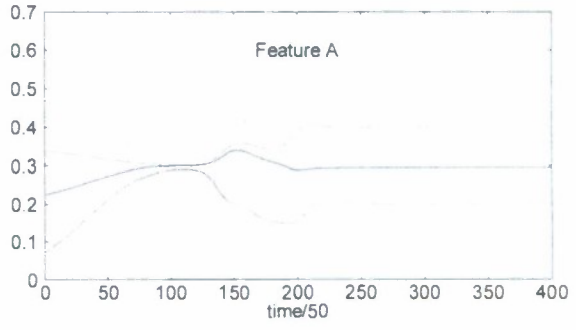

Figure 2 – Evolution of the component $e = 1$ (feature A) of the six modeling fields when the agent without language perceives the environment composed of the six "objects" illustrated in Figure 1. Notice that the agent is unable to distinguish between the two pixel blobs located at (0.29,0.29) and (0.3,0.3).

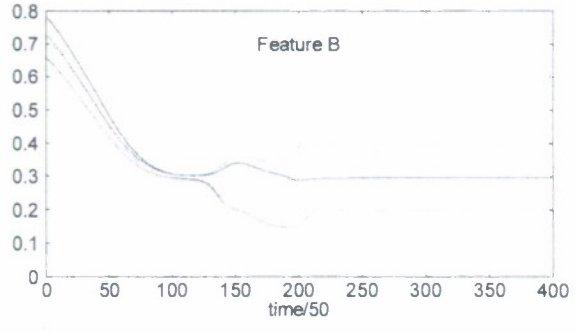

Figure 3 –Same as Figure 2 but for component $e = 2$ (feature B) of the modeling fields. So, even considering both features A and B the agent without language cannot identify all objets.

5. AGENT WITH LANGUAGE

Following Refs. [5], [6], we assume that besides the physical stimuli (O_{1i}, O_{2i}) with $i = 1,\ldots,N$ the agent receives from the environment an additional "linguistic" input W_i associated to each of the N pixels. In practice, this amounts to assume the existence of an external teacher who, while pointing to a pixel (O_{1i}, O_{2i}), utters

the word W_i. Of course, the teacher utters the same word for all pixels that make up a single object or for instances of a same category. Hence, W_i, $i=1,\ldots,N$, takes on only six different values (i.e., there are only six different words). The nature of the signals W_i (i.e., the words) is completely distinct from that of the inputs (O_{1i},O_{2i}). To take this into account we assume that the words W_i take on the integer values 1,2,…,6.

From the mathematical aspect, inclusion of the additional, linguistic component to characterize the pixels does not alter in any essential way the basic equations of the field dynamics, Eqs. (4) and (5). In particular, the inputs are now described by the triples (O_{1i},O_{2i},W_i) $i=1,\ldots,N$ which should be matched by the three-component modeling fields (S_{1k},S_{2k},S_{3k}) $k=1,\cdots,M$. Hence, the form of the field equations is unaltered, and the addition of a third component is considered by letting the index e run from 1 to 3. The parameters for the linguistic component $e=3$ are $\alpha_3=1\times10^{-4}$, $\sigma_{a3}=3$ and $\sigma_{b3}=0.1$. The reason for the larger value of σ_{a3}, as compared with the values of σ_{a1} and σ_{a2}, is that the separation between the target words are greater than the distance between the mean values of the gaussian distributions used to generate the pixels of Figure 1. We note that for the successful convergence of the MFT scheme one should always start with large fuzziness to guarantee that at the outset any one model has a nonzero similarity with all input data [9]. Moreover, the choice of a smaller value of α_3 emphasizes the need for different learning times for assimilation of inputs of distinct nature. Here we let the linguistic component evolve much slower than the non-linguistic ones. .Because of this slower rate, the time scale for convergence of the dynamics is increased as seen in the next four figures.

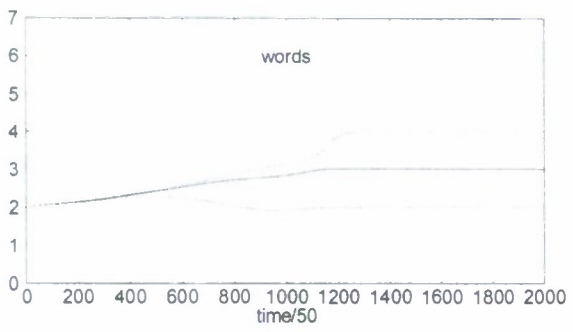

Figure 4 – Evolution of the linguistic component ($e=3$) of the modeling fields whose initial values were chosen randomly among the integers 1,2,…,6.

In Figures 4 to 7 we show the time evolution of the three components of the modeling fields in the case the linguistic input is considered.

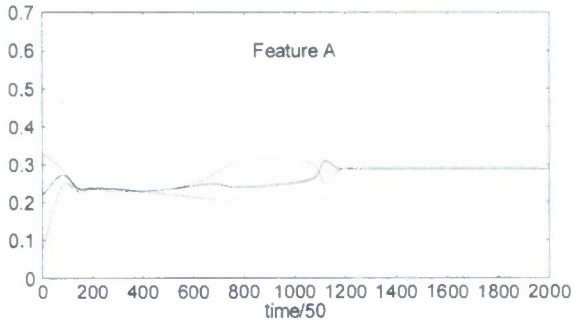

Figure 5 – Evolution of the component $e=1$ (feature A) of the six modeling fields for the agent with language. Though barely visible in this scale the agent identifies now six distinct objects or categories (see Figure 6).

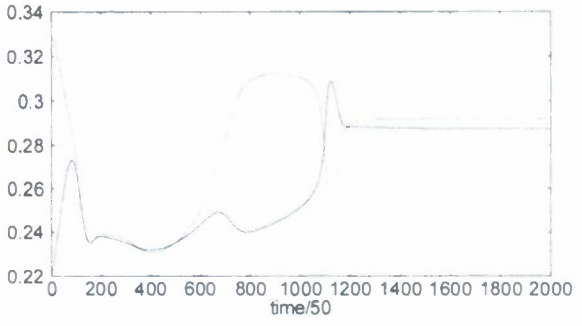

Figure 6 – A closer view of the modeling fields associated to the two overlapping objects displayed in Figure 5 confirms that the agent indeed discriminates Feature A.

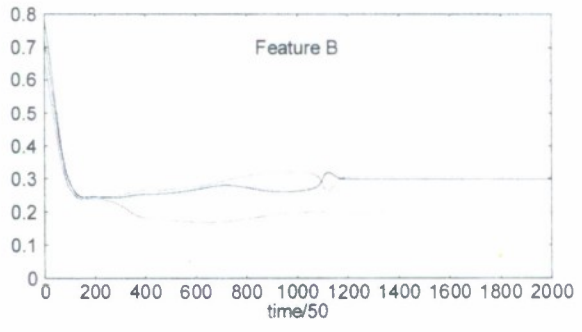

Figure 7 – Evolution of the component $e=2$ (feature B) of the six modeling fields for the agent with language. The distinction between the two overlapping objects is more perceptibe for this component.

The one-to-one correspondence between the input-words and the component $e = 3$ of the modeling fields is easily achieved as shown in Figure 4. As soon as the agent assimilates the fact that words 3 and 4 are different, which happens at $t = 50 \times 600$ approximately, the two overlapping objects are differentiated, as illustrated in Figures 5, 6, and 7. This is a remarkable finding: the extra information carried by the linguistic component allowed the agents to create distinct non-linguistic representations for the objects. We note that the asymptotic values of the modeling fields illustrated in these figures do not match exactly the means of the Gaussian distributions used to generate the pixels of Figure 1, but they are close enough to them to identify unambiguously the six objects or categories.

6. CONCLUSIONS

We have reported a computational experiment in which the addition of language, or more precisely of a linguistic signal, affects the manner that an agent processes its other sensory inputs. Remarkably, the agent with language is capable of differentiating objects or categories that it could not distinguish without language. We note that what distinguishes linguistic signals (e.g., word sounds) from other stimuli is that the agent experiences the sounds in concomitance with non-linguistic experience. The crucial role played by the linguistic signal in our experiment contrasts with the more mildly claim that language enhances performance only if the agent has already evolved an ability to respond appropriately to the visually perceived objects without language [5].

In our scenario, the agent develops only the capacity to "understand" the words uttered by the external teacher; the production of words was not considered as it must necessarily involve at least two agents (see below). The agent "understands" the meaning of a word when it associates that word stimulus with a concrete object in the environment. This type of association is made very simple in the MFT framework. In that sense, the experiment reported here is relevant to the issue of language acquisition by children [16].

As pointed above, the study of the emergence of the ability to produce linguistic signals requires the use of two or more agents. The main difficulty to adapt our discrimination task scenario to the multi-agent situation, and so replace the external teacher by the agents themselves, is that we need to assume that one agent (the speaker) can somehow distinguish between the overlapping objects while the other agent (the hearer) cannot. This type of unwarranted assumption is made in the mushroom world scenario [5], [6]. A less far-fetched possibility is to assume that the agents can perceive different features of the objects. So it is plausible to admit that what is seen as a single object by one agent is perceived as two or more objects by another agent, since they process different features of their environment. Work in this direction is on the way.

ACKNOWLEDGMENTS

Effort sponsored by the Air Force Office of Scientific Research, Air Force Material Command, USAF, under grants number FA9550-06-1-0202 and FA8655-05-1-3031. The U. S Government is authorized to reproduce and distribute reprints for Governmental purpose notwithstanding any copyright notation thereon.

REFERENCES

[1] S. Pinker and P. Bloom, "Natural languages and natural selection", *Behavioral and Brain Sciences* **13**, 707-784, 1990.

[2] W. H. Calvin and D. Bickerton, *Lingua ex Machina*, Cambridge: MIT Press, 2000.

[3] D. Bickerton, *Language & Species*, Chicago: University f Chicago Press: 1990.

[4] R. Dunbar, *Grooming, Gossip, and the Evolution of Language*, Cambridge: Harvard University Press, 1998.

[5] D. Parisi and A. Cangelosi, "A Unified Simulation Scenario for Language Development, Evolution and Historical Change," In: A. Cangelosi, D. Parisi (Eds.), *Simulating the Evolution of Language*, London: Springer-Verlag, pp. 255-275, 2002.

[6] A. Cangelosi, "Evolution of Communication and Language using Signals, Symbols and Words," *IEEE Transactions in Evolution. Computation* **5**, 93-101, 2001.

[7] S. Harnard, "The symbol grounding problem", *Physica D* **42**, 335-346, 1990.

[8] L. I. Perlovsky, *Neural Networks and Intellect: Using Model-Based Concepts*, Oxford: Oxford University Press, 2001.

[9] J. F. Fontanari and L. I. Perlovsky, "Meaning Creation and Modeling Field Theory", *Proceedings of the IEEE Conference on Integration of Knowledge Intensive Multi-Agent Systems KIMAS 05*, 405 –410, 2005.

[10] J. F. Fontanari and L.I. Perlovsky, "Categorization and symbol grounding in a complex environment", *Proceedings of the IEEE International Joint Conference on Neural Networks IJCNN 06*, 10039 –10045, 2006.

[11] L. I. Perlovsky, "Integrating language and cognition," *IEEE Connections* **2**, 8-13, 2004.

[12] L. I. Perlovsky, "Integrated Emotions, Cognition, and Language", *Proceedings of the IEEE International Joint Conference on Neural Networks IJCNN 06*, 1570 –1575, 2006.

[13] V. Cherkassky and F. Mulier, *Learning from Data: Concepts, Theory, and Methods*, New York: Wiley-Interscience, 1998.

[14] H. Akaike, "Statistical predictor identification,".Ann. Stat. Math. **22**, 203-217, 1970.

[15] D. Marr, *Vision*, San Francisco: Freeman, 1982.
[16] P. Bloom, How children learn the meaning of words, Cambridge: MIT Press, 2000.
[17]V. Tikhanoff, J.F. Fontanari, A. Cangelosi and L.I. Perlovsky, "Language and Cognition Integration Through Modeling Field Theory: Category Formation for Symbol Grounding", Lecture Notes in Computer Science vol. **4131**, 376-385, 2006.

Language acquisition and category discrimination in the Modeling Field Theory framework

José F. Fontanari, and Leonid I. Perlovsky

Abstract—**We propose a categorization task scenario to study language acquisition in which an agent receives linguistic input from an external teacher, in addition to sensory stimuli from the objects that make up the environment. The agent is endowed with the Modeling Field Theory (MFT) categorization mechanism, which enables it to identify overlapping categories from the exposition to hundreds of examples. Rather remarkably, we find that the agent with language is capable of differentiating object features that it could not distinguish without language. In this sense, the linguistic stimuli prompt the agent to redefine and refine the discrimination capacity of its sensory channels.**

I. Introduction

A major challenge to the paladins of the viewpoint that language has evolved through natural selection as any other biological organ [1] is to identify the nature of the selective pressures accountable for the emergence of language - a capability that singles out the human species from the other animals in the planet. In this contribution we examine the suggestion that such selective pressures have come from the exigencies of survival, e.g., hunting, food gathering and predator detection (see [2], [3]). We refer the reader to Ref. [4] for the alternative stance that the leading role in language evolution was played instead by the demands of the social life of early hominids. However, rather than focusing on the evolution issue, here we pursue a more pragmatic approach and consider these elementary survival needs as problems to be solved by the individuals, and ask whether and how communication can improve their performance to solve categorization problems.

The specific task we consider in this contribution is the differentiation problem, i.e., how individuals (agents) develop a more detailed knowledge of their surroundings. In fact, as pointed out by Parisi and Cangelosi [5] (see also [6]) one possible advantage of communication is that a group of individuals with this capacity can perceive their environment beyond the limits of their senses: an individual unable to communicate can access its environment based on the information provided by its own senses only. Here we use the term differentiation as synonymous to discrimination. To be able to discriminate is to be able to judge whether two inputs are the same or different. According to Harnard [7], discrimination is independent of identification as one can discriminate things without knowing what they are. For identification, the iconic representations of the raw input data must be reduced to a small set of invariant features that will reliably distinguish members of different categories. Recently we have shown how a novel adaptive approach to concept formation - Modeling Field Theory (MFT) [8] - can successfully integrate and implement these two tasks into a simple autonomous neural-networks-like scheme [9], [10]. Here we advance further and allow the agents endowed with a categorization system based on MFT to learn from an external teacher symbolic or linguistic representations for members of a category, i.e., to name the category. This is a modest first step towards the ambitious program of fully integrating language and cognition [11], [12].

In the next section, we describe the environment in which the agent lives as well as the task posed to it. In section 3 we briefly review the MFT formalism within the context of the specific categorization problem addressed in this paper. In section 4 we present the results of the simulation of the MFT dynamics in the case the agent does not receive a linguistic input, and in section 5 we address the more interesting case where an external teacher names the objects as they are perceived by the agent. Finally, section 6 summarizes the main conclusions.

II. The Categorization Task

We assume that the world contains a certain number of categories whose examples are modeled by overlapping sets of points drawn from Gaussian distributions. Figure 1 displays the particular instance we will consider in this paper. We can view each point in the figure as representing two particular features, feature A and feature B, of 600 objects (examples) which belong to six distinct categories. The issue is then to identify these categories. Note that identification presumes prior discrimination. The environment is set so that the agent cannot categorize and identify all examples, because of the considerable overlap between their features (see Fig. 1). Inspired by the

Effort sponsored by the Air Force Office of Scientific Research, Air Force Material Command, USAF, under grant number FA9550-06-1-0202. The U. S Government is authorized to reproduce and distribute reprints for Governmental purpose notwithstanding any copyright notation thereon. This research was supported also by CNPq and FAPESP, Project No. 04/06156-3.

J. F. Fontanari is with Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, SP Brazil (phone: +55-16-33739849; fax: +55-16-33739877; e-mail: fontanari@ifsc.usp.br).

L. I. Perlovsky is with Harvard University, Cambridge MA and The Air Force Research Laboratory, SN, Hanscom Air Force Base, MA (e-mail: Leonid.Perlovsky@hanscom.af.mil).

"mushroom" world scenario [5], [6], we allow the agent to receive from the environment an additional sensory input: a heard linguistic signal. Here, we assume that this signal is produced by an external teacher who has perfect knowledge of the agent's environment. In the following we will show that when the agent receives the linguistic input associated to the different objects, it can create new concept-models and so identify unambiguously all objects.

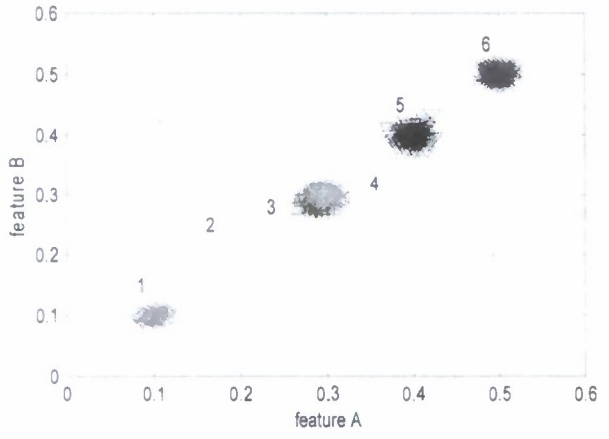

Fig. 1. The six sets of 100 examples, represented by the features A and B (e.g., texture and color), of six categories. The coordinates of each pixel are drawn from Gaussian distributions of means 0.1, 0.2, 0.29, 0.3, 0.4, and 0.5 and standard deviation 0.01. The labels 1, 2, ..., 6 are the words (names) associated to each example of the six categories.

III. The Modeling Field Theory approach

The basic idea behind Modeling Field Theory is the association between lower-level signals (e.g., inputs) and higher-level concept-models (internal representations) avoiding the combinatorial complexity inherent to such a task. This is achieved by using measures of similarity between concept-models and input signals together with a new type of logic, so-called fuzzy dynamic logic. We refer the reader to Ref. [8] for a complete presentation of MFT; here we particularize the general framework to the problem of categorizing the $N = 600$ examples of the six categories depicted in Fig. 1. Each example is described by the pair of real variables (O_{1i}, O_{2i}) with $i = 1, \ldots, N$. Let us assume that there are M concept-models described by the pairs (S_{1k}, S_{2k}) with $k = 1, \cdots, M$ that should "model" (i.e., create iconic representations) the original examples. Hence the denomination "modeling fields" to the mathematical quantities S_{ek}. We define arbitrarily the following partial similarity measure between object i and concept k

$$l(i \mid k) = \prod_{e=1}^{2} \left(2\pi\sigma_{ek}^2\right)^{-1/2} \exp\left[-(O_{ei} - S_{ek})^2 / 2\sigma_{ek}^2\right] \qquad (1)$$

where, at this stage, the fuzziness σ_{ek} are parameters given *a priori*. We refer the reader to Ref. [13] for another application of MFT in the case of multi-component inputs. The goal is to find an assignment between models and examples such that the global similarity

$$L = \prod_i \sum_k l(i \mid k) \qquad (2)$$

is maximized. The maximization of L can be achieved using the MFT mechanism of concept formation which is obtained through the direct maximization of (2) with respect to S_{ek}. The aim here is to derive a dynamical equation for the modeling fields S_{ek} such that $dL/dt \geq 0$ for all time t. This condition can easily be met by choosing $dS_{ek}/dt = \partial L/\partial S_{ek}$ since then

$$dL/dt = \sum_{e,k} (\partial L/\partial S_{ek})(dS_{ek}/dt) = \sum_{e,k} (\partial L/\partial S_{ek})^2 \geq 0 \qquad (3)$$

as required. The calculation of $\partial L/\partial S_{ek}$ is straightforward

$$\frac{\partial L}{\partial S_{ek}} = \sum_i \frac{1}{\sum_{k'} l(i \mid k')} \frac{\partial l(i \mid k)}{\partial S_{ek}} \qquad (4)$$

and leads to the following dynamics for the modeling fields

$$dS_{ek}/dt = \sum_i f(k \mid i)\left[\partial \log l(i \mid k)/\partial S_{ek}\right], \qquad (5)$$

for $e = 1,2$ and $k = 1, \cdots, M$ and where we have used the identity $\partial y/\partial x = y\,\partial \log y/\partial x$. The fuzzy association variables $f(k \mid i)$ defined by

$$f(k \mid i) = l(i \mid k) \Big/ \sum_{k'} l(i \mid k') \qquad (6)$$

play a fundamental role in the interpretation of the MFT dynamics by giving a measure of the correspondence between object i and concept k relative to all other concepts k'. We note that although the features A and B of an example are independent random variables, the two components of a modeling field are coupled dynamic variables. Actually, the term $f(k \mid i)$ in Eq. (5) couples not only S_{1k} and S_{2k} but also components of different modeling fields [13].

It can be shown that the dynamics (5) always converges to a (possibly local) maximum of the similarity L [8]. By properly adjusting the fuzziness σ_{ek} the global maximum can be attained. A salient feature of dynamic logic is a match between parameter uncertainty and fuzziness of similarity. In what follows we decrease the fuzziness during the time evolution of the modeling fields according to the following prescription

$$\sigma_{ek}^2(t) = \sigma_{ae}^2 \exp(-\alpha_e t) + \sigma_{be}^2 \qquad (7)$$

with $\alpha_e = 5\times10^{-4}$, $\sigma_{be} = 0.03$, and $\sigma_{ae} = 0.5$ for $e = 1,2$. Note that these parameters are the same for all models $k = 1,\ldots,M$ and components $e = 1,2$. However, we will need different parameter values to stabilize the linguistic component $e = 3$, which we will introduce in the Section 5. As a guideline for setting the parameter values in (7) we note that σ_{ae} must be chosen large enough such that, at the beginning, all examples are described by all fields, whereas the baseline resolution σ_{be} must be small enough such that, at the end, a given field will describe a single category. However, σ_{be} should not be set to a too small value to avoid numerical instabilities in the calculation of the partial similarities (1).

A word is in order about the connection between the MFT framework and neural networks. A MFT neural architecture was described in [8], which combines architecture with models of objects or category examples. Essentially, input neurons or bottom-up signals encode the feature values of the category examples O_{ei}, and top-down or priming signal-fields to these neurons are generated by the modeling fields S_{ek}. Interaction between bottom-up and top-down signals is determined by the neural weights $f(k|i)$ that associate signals and models. As described before, these weights are functions of the model parameters S_{ek}, which in turn are dynamically adjusted so as to maximize the overall similarity between category examples and models. This formulation sets MFT apart from many other neural networks. There is, on the other hand, a certain formal similarity between the MFT approach and the Hopfield-Tank neural network approach to tackle optimization problems [14]. This becomes apparent when one recognizes that the nature of perceptual problems dealt with here is similar to that of other optimization problems. In fact, in both systems it is the time evolution of analog neurons that drives the neural configuration to a maximum of the cost function [the global similarity (2) in our case]. Moreover, the quality of the solutions found by the neural network is greatly improved by annealing the analog gain parameter [15], in a similar manner as the slow decrease of the fuzziness according to (7) leads ultimately to perfect categorization. In addition, the competition between different concept-model to match the category examples is reminiscent of the dynamics of unsupervised learning algorithms and, in particular, of self-organizing maps [16].

In what follows we will set $M = 6$ so Eq. (5) stands for a set of twelve nonlinear coupled equations, which are solved with Euler's method using the step-size $h = 10^{-5}$. In a previous contribution [10], we have combined the MFT approach with the Akaike Information Criterion [17] to design a categorization system that infers correctly the true number of objects in an environment similar to that exhibited in Fig. 1, but for the purposes of this paper, any choice of $M \geq 6$ will be satisfactory.

IV. AGENT WITHOUT LANGUAGE

The choice of the particular environment depicted in Fig. 1 is motivated by the inability of the agent to distinguish between the six categories into which the 600 examples are naturally organized. The difficulty, of course, is to distinguish between the two sets of examples centered at the coordinates (0.29,0.29) and (0.3,0.3), which are labeled 3 and 4, respectively, by the external teacher. Figs. 2 and 3 illustrate the time dependence of the two components, $e = 1$ (feature A) and $e = 2$ (feature B) of the modeling fields S_{ek}. Since feature B is essentially equivalent to feature A the associated modeling field components exhibit a very similar behavior pattern. What distinguish these components are simply their initial values, which were chosen randomly from the uniform distribution. The point of Figs. 2 and 3 is to stress the rather expected failure of most categorization methods to distinguish highly overlapping categories. The agent is able to identify four of the six categories displayed in Fig. 1 and, in addition, it can clearly discriminate the two overlapping categories from the other four.

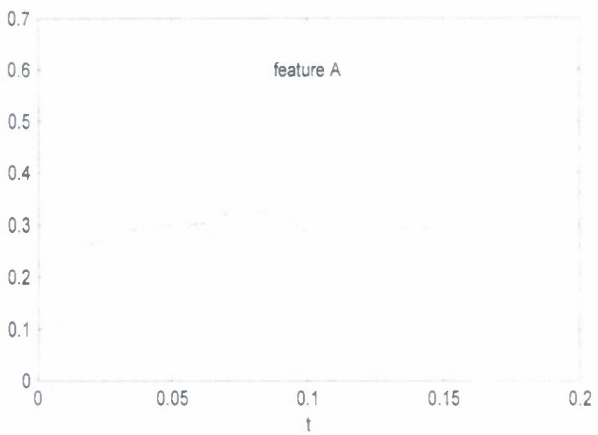

Fig. 2. Evolution of the component $e = 1$ (feature A) of the six modeling fields when the agent without language perceives the environment composed of 600 examples that belong to six categories as illustrated in Fig. 1. Note that the agent is unable to distinguish between the two categories labeled 3 and 4 by the external teacher

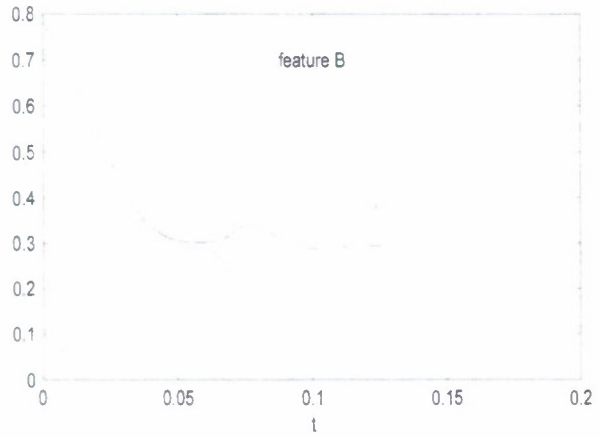

Fig. 3. Same as Fig. 2, but for component $e = 2$ (feature B) of the modeling fields. So, even considering both features A and B the agent without language cannot identify all six categories.

V. AGENT WITH LANGUAGE

Motivated by the "mushroom world" scenario [5], [6], we assume that besides the physical stimuli (O_{1i}, O_{2i}) with $i = 1,\ldots,N$ the agent receives from the environment an additional "linguistic" input W_i associated to each of the $N = 600$ examples depicted in Fig. 1. In practice, this amounts to assume the presence of an external teacher who, while pointing to an example (O_{1i}, O_{2i}), utters the word W_i. Of course, the teacher utters the same word for all examples belonging to a same category. Hence, W_i, $i = 1,\ldots,N$, takes on only six different values (i.e., there are only six different words). The nature of the signals W_i (i.e., the words) is completely distinct from that of the inputs (O_{1i}, O_{2i}). To take this into account we assume that the words W_i take on the integer values $1,2,\ldots,6$, rather than real values as the modeling fields associated to the physical features A and B. These are the category labels exhibited in Fig. 1.

From the mathematical aspect, inclusion of the additional, linguistic component to characterize the examples does not alter in any essential way the basic equations of the field dynamics, Eqs. (5) - (7). In particular, the inputs are now described by the triples (O_{1i}, O_{2i}, W_i) $i = 1,\ldots,N$ which should be matched by the three-component modeling fields (S_{1k}, S_{2k}, S_{3k}) $k = 1,\cdots,M$. Hence, the form of the field equations is unaltered, and the addition of a third component is taken into account by letting the index e run from 1 to 3. The parameters for the linguistic component $e = 3$ are $\alpha_3 = 1\times10^{-4}$, $\sigma_{a3} = 3$ and $\sigma_{b3} = 0.1$. The reason for the larger value of σ_{a3}, as compared with the values of σ_{a1} and σ_{a2}, is that the separation between the target words are greater than the distance between the mean values of the gaussian distributions used to generate the category examples in Fig. 1. We note that for the successful convergence of the MFT scheme one should always start with large fuzziness to guarantee that at the outset all models have a nonzero similarity with all input data [9]. Moreover, the small magnitude of α_3 as compared with α_1 and α_2 emphasizes the need for different learning times for assimilation of inputs of distinct nature. Here we let the linguistic component evolve much slower than the non-linguistic ones. Because of this slower rate, the time scale for convergence of the dynamics is increased by a factor 5 as seen in the next four figures.

In Figs. 4 to 7 we show the time evolution of the three components of the modeling fields in the case the linguistic input is considered.

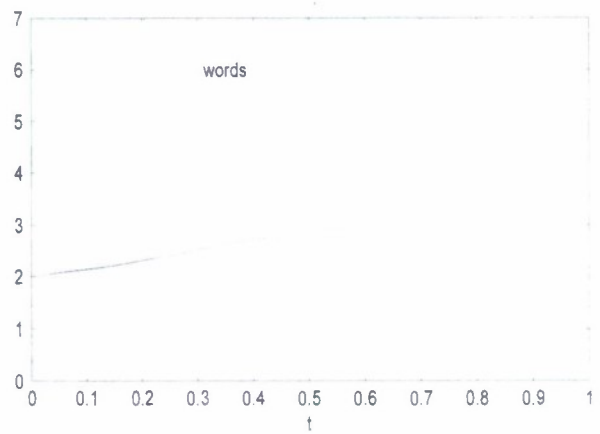

Fig. 4. Evolution of the linguistic component ($e = 3$) of the modeling fields whose initial values were chosen randomly among the integers $1,2,\ldots,6$.

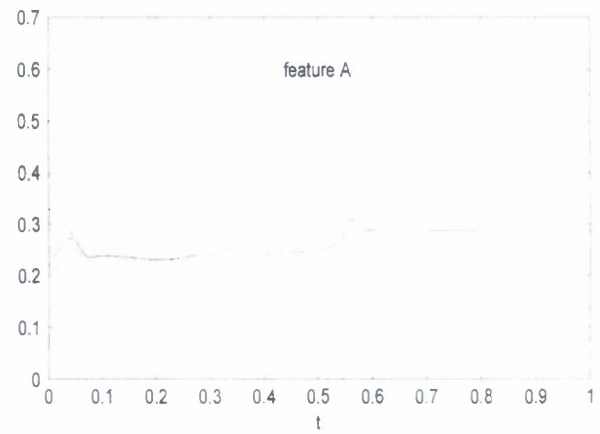

Fig. 5. Evolution of the component $e = 1$ (feature A) of the six modeling fields for the agent with language. Though barely visible in this scale the agent identifies the six distinct objects or categories (see Fig. 6).

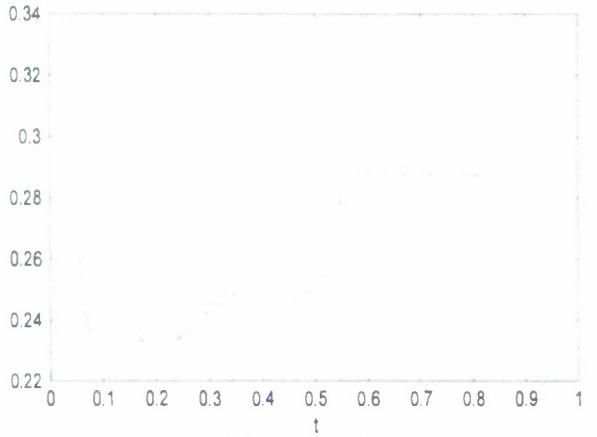

Fig. 6. A closer view of the modeling fields associated to the two overlapping objects displayed in Fig. 5 confirms that the agent indeed discriminates feature A.

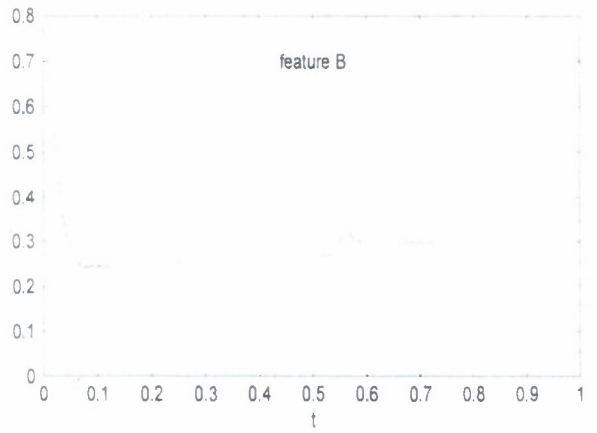

Fig. 7. Evolution of the component $e = 2$ (feature B) of the six modeling fields for the agent with language. The distinction between the two overlapping objects is more perceptible for this component.

The one-to-one correspondence between the input-words and the component $e = 3$ of the modeling fields is easily achieved as shown in Fig. 4. As soon as the agent assimilates the fact that words 3 and 4 are different, which happens at $t = 0.3$ approximately, the two overlapping objects are differentiated, as illustrated in Figs. 5, 6, and 7. This is a remarkable finding: the extra information carried by the linguistic component allowed the agent to create distinct non-linguistic representations for the category examples. In other words, the knowledgement that the categories 3 and 4 are distinct, that was achived through the linguistic input, allowed the agent to redefine and refine its expectations about features A and B. We wonder whether the agent would create ficticious distinctions between those features in the case their distributions were identical. As already said, what is behind this result is the inter-dependence of the features introduced in the field equations by the fuzzy association variables, so that the learning of one of the features affects the learning of the others Finally, we note that the asymptotic values of the modeling fields illustrated in these figures do not match exactly the means of the Gaussian distributions used to generate the data of Fig. 1, but they are close enough to them to identify unambiguously the six categories. We have verified that the matching improves when the number of examples of each category increases.

VI. Conclusion

We have reported a computational experiment in which the addition of language, or more precisely of a linguistic signal, affects the manner that an agent processes its other sensory inputs. Remarkably, the agent with language is capable of differentiating categories that it could not distinguish without language. We note that what distinguishes linguistic signals (e.g., word sounds) from other stimuli is that the agent experiences the sounds in concomitance with non-linguistic experience. The crucial role played by the linguistic signal in our experiment contrasts with the more mildly claim that language enhances performance only if the agent has already evolved an ability to respond appropriately to the visually perceived objects without language [5].

In our scenario, the agent develops only the capacity to "understand" the words uttered by the external teacher; the production of words was not considered as it must necessarily involve at least two agents (see below). The agent "understands" the meaning of a word when it associates (i.e., grounds) that word stimulus with a concrete object or category example in the environment [18]. This type of association is made very simple in the MFT framework. Since in the experiment reported here it is assumed the presence of an external teacher with complete knowledge of the agent's environment our results are relevant to the issue of language acquisition by children (see, e.g., [19]) rather than to the language evolution problem.

As pointed above, the study of the emergence of the ability to produce linguistic signals requires the use of two or more agents. There are two obstacles to adapt our discrimination task scenario to the multi-agent situation, and so replace the external teacher by the agents themselves. First, we need to assume that one agent (the speaker) can somehow distinguish between the overlapping categories while the other agent (the hearer) cannot. This type of unwarranted assumption is made in the mushroom world scenario [5], [6]. A less far-fetched possibility is to assume that the agents can perceive different features of the examples. So it is plausible to admit that examples put into a same category by one agent are perceived as completely distinct by another agent, because they process different features of their environment. Second, the MFT framework suits well to aprehend characteristics of the environment that are produced by a well defined process that may or may not be corrupted by noise. However, the mechanism that leads two agents to reach a consensus on which word to assign to a given category is not described by such a process but by a series of

guessing or naming games (see, e.g., [20], [21]) in which similarity measures seem to play no role at all. In fact, consider the case where two agents assign different words to the same category and each agent broadcasts its linguistic signal to the other. There is no reason for an agent to give up its word in favor of that used by the other agent (as actually happens in the case of an external teacher considered here) and so no consensus will ever be reached in this case. Hence, whereas the present framework looks very promising to model acquisition of language in children, we do not see how it could be applied to the formation of a common lexicon in a community of agents. We refer the reader to Ref. [22] for a framework that uses MFT to categorize examples and the guessing games strategy to name the just created categories.

REFERENCES

[1] S. Pinker and P. Bloom, "Natural languages and natural selection", *Behavioral and Broin Sciences*, vol. 13, pp. 707-784, 1990.

[2] W. H. Calvin and D. Bickerton, *Lingua ex Machina*, Cambridge: MIT Press, 2000.

[3] D. Bickerton, *Language & Species*, Chicago: University f Chicago Press: 1990.

[4] R. Dunbar, *Graoming, Gossip, ond the Evolution of Longuage*, Cambridge: Harvard University Press, 1998.

[5] D. Parisi and A. Cangelosi, "A Unified Simulation Scenario for Language Development, Evolution and Historical Change," In: A. Cangelosi, D. Parisi (Eds.), *Simulating the Evolution of Language*, London: Springer-Verlag, pp. 255-275, 2002.

[6] A. Cangelosi, "Evolution of Communication and Language using Signals, Symbols and Words," *IEEE Transactions in Evalution. Computation*, vol. 5, pp. 93-101, 2001.

[7] S. Harnard, "The symbol grounding problem", *Physica D* , vol. 42, pp. 335-346, 1990.

[8] L. I. Perlovsky, *Neural Netwarks and Intellect: Using Model-Bosed Concepts*, Oxford: Oxford University Press, 2001. J. F. Fontanari and L. I. Perlovsky, "Meaning Creation and Modeling Field Theory", *Proceedings of the IEEE Conference on Integrotion of Knowledge Intensive Multi-Agent Systems KIMAS 05*, pp. 405 –410, 2005.

[9] J. F. Fontanari and L.I. Perlovsky, "Categorization and symbol grounding in a complex environment", *Proceedings of the IEEE Internotionol Joint Conference an Neural Netwarks IJCNN 06*, pp. 10039 –10045, 2006.

[10] L. I. Perlovsky, "Integrating language and cognition," *IEEE Connections*, vol. 2, pp. 8-13, 2004.

[11] L. I. Perlovsky, "Integrated Emotions, Cognition, and Language", *Proceedings of the IEEE Internotionol Jaint Conference on Neural Networks IJCNN 06*, pp. 1570 –1575, 2006.

[12] V. Tikhanoff, J.F. Fontanari, A. Cangelosi and L.I. Perlovsky, "Language and Cognition Integration Through Modeling Field Theory: Category Formation for Symbol Grounding", *Lecture Notes in Computer Science*, vol. 4131, pp. 376-385, 2006.

[13] J. J. Hopfield and D. W. Tank, "Neural computation of decisions in optimization problems," *Biological Cybernetics*, vol. 52, pp. 141-152, 1985.

[14] D. E. Vandenbout and T. K. Miller, "Improving the performance of the Hopfield-Tank neural network through normalization and annealing," *Bialogical Cybernetics*, vol. 62, pp. 129-139, 1989.

[15] T. Kohonen, *Self-Organizing Mops*, Springer Series in Information Sciences, vol. 30, 1995.

[16] H. Akaike, "A new look at the statistical model identification", *IEEE Tronsoctions on Automatic Control*, vol. 19, pp. 716-723, 1974.

[17] A. Cangelosi, A. Greco, and S. Harnard, "Symbol grounding and the symbolic theft hypothesis," In: A. Cangelosi, D. Parisi (Eds.), *Simulating the Evalution of Language*, London: Springer-Verlag, pp. 191-210, 2002.

[18] P. Bloom, How children learn the meaning of words, Cambridge: MIT Press, 2000.

[19] L. Steels, and F. Kaplan, "Spontaneous lexicon change", *Proceedings of COLING-ACL*. Morgan Kaufmann, San Francisco, pp. 1243-1250, 1998.

[20] J. de Beule, B. de Vylder, and T. Belpaeme, "A cross-situational learning algorithm for damping homonymy in the guessing game," In: L.M. Rocha, M. Bedau, D. Floreano, R. Goldstone, A. Vespignani, L. Yaeger (Eds.), *Proceedings of the Xth Conference on Artificial Life*, The MIT Press, Cambridge, 2006.

[21] J. F. Fontanari and L. I. Perlovsky, "Meaning creation and communication in a community of agents", *Proceedings of the IEEE Internationol Jaint Conference on Neurol Netwarks IJCNN'06*, pp. 2892 –2897, 2006.

Integrating Language and Cognition: A Cognitive Robotics Approach

Angelo Cangelosi (1), Vadim Tikhanoff (1), José Fernando Fontanari (2), and Emmanouil Hourdakis (3)

(1) Adaptive Behaviour and Cognition Research Group, University of Plymouth, Drake Circus, Plymouth PL4 8AA, UK (acangelosi@plymouth.ac.uk)
(2) Instituto de Física de São Carlos, Universidade de São Paulo, 13560-970 São Carlos SP, Brazil
(3) Institute of Computer Science, University of Crete, Greece

Abstract— **In this paper we present some recent cognitive robotics studies on language and cognition integration to demonstrate how the language acquired by robotic agents can be directly grounded in action representations. These studies are characterized by the hypothesis that symbols are directly grounded into the agents' own categorical representations, whilst at the same time having logical (e.g. syntactic) relationships with other symbols. The two robotics studies are based on the combination of cognitive robotics with neural modeling methodologies such as connectionist models and modeling field theory. Simulations demonstrate the efficacy of the mechanisms of action grounding of language and the symbol grounding transfer in agents that acquire a lexicon via imitation and linguistic instructions. The paper also discusses the scientific and technological implications of such an approach.**

I. Introduction

Recent advances in cognitive psychology, neuroscience and linguistics support an embodied view of cognition, i.e. the fact that cognitive functions (perception, categorization, reasoning, language) are strictly intertwined with sensorimotor and emotional processes (Wilson 2002). This is particularly evident in recent studies on the grounding of language in action and perception (Pecher & Zwann 2004). For example, in psycholinguistics, Glenberg & Kaschak 2002 have demonstrated the existence of Action-sentence Compatibility Effects. In sentence comprehension tasks, participants are faster to judge the sensibility of sentences implying motion toward the body (e.g. "Courtney gave you the notebook") when the response requires moving toward the body (i.e. press a button nearer body). When the sentence implied movement away from the body, participants were faster to respond by literally moving away from their bodies (press a button farther from body). The data support an embodied theory of meaning that relates the meaning of sentences to human action and motor affordances. This is also consistent with neuroscientific studies on action and language, such as the involvement of the mirror neuron system for action and language learning (Rizzolatti & Arbib 1998) and brain imaging studies where words (e.g. action verbs) activate cortical areas (e.g. motor and premotor cortex) in a somatotopic fashion (Pulvermuller 1993). In linguistics, the link between the properties of language and their relationship with cognitive processes has been formalized by cognitive and constructivist linguistic theories (e.g. Talmy, 1980).

This growing empirical evidence is consistent with recent advances in artificial intelligence and robotics, where the design of the capabilities of the artificial cognitive agents is based on an integrated cognitive approach (Perlovsky, this volume). For example, the design of the linguistic capabilities of interactive systems for human-robot communication are built (grounded) onto the robot's other

This work was in part sponsored by the Air Force Office of Scientific Research, Air Force Material Command, USAF, under grants number FA8655-05-1-3060 and FA9550-06-1-0202. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purpose notwithstanding any copyright notation thereon.

sensorimotor and cognitive skills (Cangelosi et al. 2005; Feldman & Narayanan 2004). Robots acquire words through direct interaction with their physical and social world, so that linguistic symbols do not exist as arbitrary representations of some notion, but are intrinsically connected to behavioral or cognitive abilities, based on the properties of the reference system they belong to. This task of connecting the arbitrary symbols used in internal reasoning with external physical stimuli is known as Symbol Grounding (Harnad 1990).

In this paper we will present some recent cognitive robotics studies on language and cognition integration to demonstrate how the language acquired by robotic agents can be directly grounded in action representations. These studies are characterized by the hypothesis that symbols are directly grounded into the agents' own categorical representations, whilst at the same time having logical (e.g. syntactic) relationships with other symbols. First, each symbol is directly grounded into internal categorical representations. These representations include perceptual categories (e.g. the concept of blue color, square shape, and male face), sensorimotor categories (e.g. the action concept of grasping, pushing, and carrying), social representations (e.g. individuals, groups and relationships) and other categorizations of the agent's own internal states (e.g. emotions and motivations). These categories are connected to the external world through our perceptual, motor and cognitive interactions with the environment. Second, symbols also have logical (e.g. syntactic) relationships with the other symbols of the lexicons used for communication. This allows symbols to be combined, using compositional rules such as grammar, to form new meanings. For example, the combination of the two symbols "stripes" and "horse", which are directly grounded into the agent's own sensorimotor experience of striped objects and horses in its environment, produces the new concept (and word) "zebra". This new symbol becomes indirectly grounded in the agents' experience of the world through the process of "symbol grounding transfer". An example of symbol grounding transfer will be demonstrated in the cognitive robotics model for the acquisition and combination of names of actions.

The two cognitive robotics models presented below will demonstrate the mechanisms of action grounding of language and the symbol grounding transfer in agents that acquire a lexicon via imitation and linguistic instructions. These models are based on the combination of cognitive robotics with neural modeling methodologies such as connectionist models and modeling field theory.

II. Cognitive Robotics and Connectionist Modelling of Symbol Grounding Transfer

Neural networks have been proposed as an ideal cognitive modeling methodology to deal with the symbol grounding problem (Harnad 1990). For example, connectionist models, such as multi-layer perceptrons (MLP), permit a good implementation of the process of grounding output symbolic representations in the (analogical) input representation of external stimuli (Plunkett et al. 1992; Cangelosi 2005). The same feedforward models can be extended to simulate the process of grounding transfer (Cangelosi et al. 2000). More recently, these connectionist models have been incorporated in studies based on cognitive agents and robots. Cognitive robotics refers to the field of robotics that aims at builds autonomous cognitive systems capable of performing cognitive tasks such as perception, categorization, language and sensorimotor problems. Cognitive robotics approaches include epigenetic robotics and autonomous mental development systems (Weng et al. 2001), as well as evolutionary robotics (Nolfi & Floreano 2000). Here we briefly present a cognitive robotics model for the acquisition of a lexicon of words of action and for the grounding transfer. This is an extension of the first cognitive robotics model for symbol grounding in language comprehension tasks originally developed by Cangelosi and Riga (2006). The new model presented below extends the previous study by considering both linguistic comprehension and production capabilities.

A. The Robot

The robotics model consists of two simulated agents (teacher and learner) embedded within a virtual simulated environment (Fig. 1). Each robot consists of two 3-segment arms attached to a torso (6 Degrees of Freedom). This is further connected to a base with four wheels, which were not used in the present simulation. Through the two arms the robot can interact with the environment and manipulate objects placed in front of it. Three objects were used in the current simulation: a cube, a horizontal plane and a vertical bar. The agent can receive in the input retina different views (perspectives) of each object. The agent has to learn six basic actions: lower right shoulder, lower left shoulder, close right upperarm, close left upperarm, close right elbow, close left elbow. They will also learn the name of such basic actions: "LOWER_RIGHT_SHOULDER", "LOWER_LEFT_SHOULDER", "CLOSE_RIGHT_UPPERARM", "CLOSE_LEFT_UPPERARM", "CLOSE_RIGHT_ELBOW", "CLOSE_LEFT_ELBOW" Each action will be associated with some of the above objects that are put in front of the agent. The close left and close right shoulder actions are associated with different views of the cube.

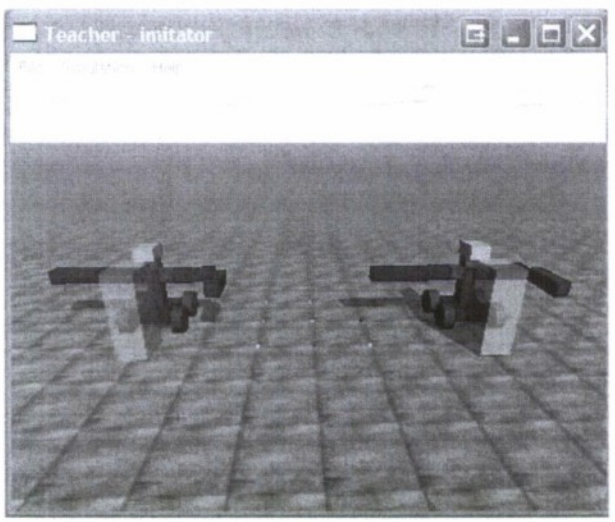

Fig. 1: Simulation setup with the two robots. The teacher robot is on the left and the learner on the right. The agents are performing the close left elbow action.

This system is implemented using ODE (Open Dynamics Engine, www.ode.org), an open source, high performance library for simulating rigid body dynamics. ODE is useful for simulating vehicles, objects in virtual reality environments and virtual creatures, and it is being increasingly used for simulation studies on autonomous cognitive systems.

The first agent, the teacher, is pre-programmed to perform and demonstrate a variety of basic actions, each associated with a linguistic signal. These are demonstrated to the second robot, the learner, which attempts to reproduce the actions by mimicking them. First the agent acquires basic actions by observing the teacher, and then it learns the basic action names (direct grounding). Subsequently, it autonomously uses the linguistic symbols that were grounded in the previous learning stage to acquire new higher-order actions (symbol grounding transfer).

B. Neural network controller and training procedure

The imitator robot is endowed with a MLP neural network (Fig. 2) with input units for vision, proprioceptive and linguistic input and output units for motor control and linguistic output. For the robot motor control, the motor output units encode the force that is being applied on each joint. Each action

consists of a sequence of 10 steps of motor activations.

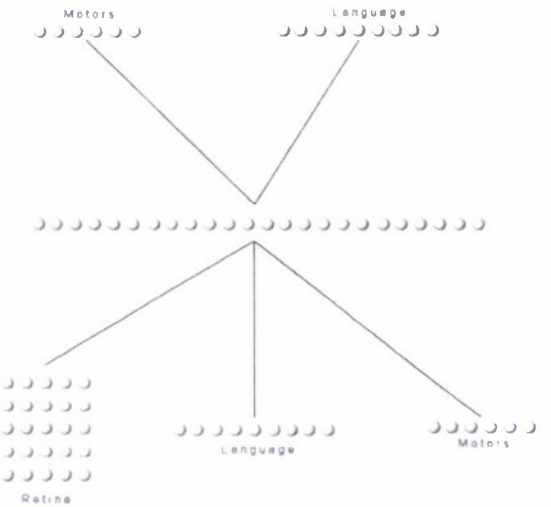

Fig 2. Architecture of the learner robot's neural network controller.

We attain the grounding transfer, using a 3 stage training process: (1) BA Basic Action learning, (2) EL Entry-Level naming and (3) HL Higher-Order learning.

During the Basic Action learning stage, the agent learns to execute all six basic actions in association with the view of the different objects. No linguistic elements are used at this stage. The imitation algorithm is used to adjust the weights contributing to the activation of the motor units using supervised learning (see Cangelosi et al. 2006 for the learning algorithm details).

The second learning stage, Entry Level naming (EL), was concerned with associating the previously acquired behaviors to linguistic signals. It features three sequential activation cycles. The first EL cycle, Linguistic Production, trains the learner how to name the 6 basic actions. Motor (proprioceptive) and visual (object view) information are given in input to the network. The agents learn to correctly activate the output linguistic nodes corresponding to the basic action names. This is based on a standard backpropagation algorithm. This linguistic production cycle implements the process of basic symbol grounding, by which the names (symbols) that the agent is learning are directly grounded on its own perceptual and sensorimotor experience. In the second EL cycle, Linguistic Comprehension, learner agents are taught to correctly respond to a linguistic signal consisting of the name of the action, without having the ability to perceive the object associated to the action. To accomplish this, the retinal units in the network were set to 0, whilst we activate the input units corresponding to the action name. In the final EL cycle, Imitation, both motor and linguistic inputs were activated in input, and the network learns to reproduce the action in output and activate the corresponding action name unit. This third cycle is necessary to permit the linking of the production and the comprehension tasks in the hidden units activation pattern (Cangelosi et al. 2000).

The final training stage, Higher-Level (HL) learning, allows the learner agents to autonomously acquire higher-order actions without the need of a demonstration from the teacher. This is achieved only through a linguistic instruction strategy and a "mental simulation" strategy similar to Barsalou's perceptual symbol system hypothesis (Barsalou 1999). The teacher has only to provide new linguistic instructions consisting of the names of two basic actions and the name of a new higher-order action. For example, the three higher-order actions, "LOWER_RIGHT_SHOULDER+LOWER_LEFT_SHOULDER=PLACE".

Once the teacher (or a human instructor) provides a higher-order instruction, the learner goes through four HL learning cycles. First it activates only the input unit of the first basic action name to produce and store ("memorize") the corresponding sequence of 10 motor activation steps. Second, it activates in input the linguistic units for the first basic action name and the new higher-order action. The resulting 10 motor activations are compared with the previously stored values to calculate the error and apply the backpropagation weight corrections. The next two cycles are the same as the first two, except that the second basic action name unit is activated as well.

The Higher-Order stage permits the implementation of a purely autonomous way to acquire new actions through the linguistic combination of previously-learned basic action names. The role of the teacher in this stage is only that of providing a linguistic instruction, without the need to give a perceptual demonstration of the new action. The motor imitation learning, such as in the Basic Action training stage, is a slow process based on continuous supervision, trial-and-error feedback-based learning. The acquisition of a new concept through linguistic instruction is, instead, a quicker learning mechanism because it requires the contribution of fewer units (the localist linguistic units) and corresponding weights. Moreover, in a related symbol grounding model on language (symbolic) vs. error-and-trial (sensorimotor toil) learning of categories, the linguistic procedure consistently outperforms the other learning method (Cangelosi & Harnad 2000).

To establish if the agent has actually learned the new high-order actions and transferred the grounding from basic action names to higher order names, a test phase is performed. This grounding transfer test aims at evaluating the aptitude of the imitator agent to perform a new composite action with any of the objects previously associated, in the absence of the linguistic descriptions of the basic actions. Thus the agent is requested to respond solely on the signal of the composite action (e.g. Grab) and selectively to the different view of the objects. In addition, while the imitator was taught only the motion of the dissected action for each composite behavior, the test evaluated the performance of the higher-order composite action. This was a behavior never seen before by the robot. The stage was comprised of two basic trials per behavior, using the different view of the objects. All inputs were propagated through the network with no training occurring.

C. Results

We replicated the simulation experiment as above with five agents. Each agent had a different set of random weights initialized in the range ±1. The three learning stage, Basic Action, Entry-Level and Higher-Level learning, respectively lasted for 1000, 3000, 1500 epochs. This was the approximate minimum number of epochs necessary to reach a good learning performance. The parameters of the backpropagation algorithm were set as follow: BA stage, momentum 0.6 and learning rate 0.2; EL stage, momentum 0.6 and learning rate 0.3; HL stage, momentum 0.8 and learning rate 0.2 . The weights were updated at the end of every action.

Overall, results indicate that all agents were able to learn successfully the 6 basic actions and the 3 higher-order behaviors. At the end of the stage, the imitator was able to execute all actions flawlessly, when presented with an object (final error of 0.004). The overall average error on the final epoch of the Entry-Level stage was 0.03. Finally, in the grounding transfer test the agent was requested to perform a new composite action by giving in input only the new action name or the new name together with the basic action names (error 0.018). These results confirm our hypothesis that previously grounded symbols are transferred to the new behaviors.

III. Action and Lexicon Scaling up with Modeling Field Theory

In this study we aim at extending the behavioral and linguistic capabilities of the robot by scaling up its action repertoire. Perlovsky (2001; this volume) has recently proposed the use of the Modeling Field

Theory (MFT) learning algorithm to deal with the issue of the combinatorial complexity (CC) of linguistic and cognitive modeling based on machine learning techniques such as multi-layer perceptrons. The Modeling Field Theory (MFT) algorithm uses dynamic logic to avoid CC and computes similarity measures between internal concept-models and the perceptual and linguistic signals. By using concept-models with multiple sensorimotor modalities, a MFT system can integrate language-specific signals with other internal cognitive representations. Perlovsky's proposal to apply MFT in the language domain is highly consistent with the grounded approach to language modeling discussed above. That is, both accounts are based on the strict integration of language and cognition. This permits the design of cognitive systems that are truly able to "understand" the meaning of words being used by autonomously linking the linguistic signals to the internal concept-models of the word constructed during the sensorimotor interaction with the environment. The combination of MFT systems with grounded agent simulations will permit the overcoming of the CC problems currently faced in grounded agent models and scale up the lexicons in terms of high number of lexical entries and syntactic categories.

Modeling Field Theory is based on the principle of associating lower-level signals (e.g., inputs, bottom-up signals) with higher-level concept-models (e.g. internal representations, categories/concepts, top-down signals) avoiding the combinatorial complexity inherent to such a task. This is achieved by using measures of similarity between concept-models and input signals together with a new type of logic, so-called dynamic logic. MFT may be viewed as an unsupervised learning algorithm whereby a series of concept-models adapt to the features of the input stimuli via gradual adjustment dependent on the fuzzy similarity measures.

A. Extended action and lexicon repertoire

The robotic scenario is based on the same simulated robotic agents described in the previous section (see Fig. 1). The teacher robot is pre-programmed to demonstrate an extended action repertoire of 112 actions. The learner robot uses MFT to learn to reproduce those actions as well as to learn the actions names.

The main difference with respect to the previous model is the use of 112 different actions. These are inspired by the semaphore flag signaling alphabet. For the encoding of the actions, we collected data on the posture of the teacher robots using 6 features, i.e. 3 pairs of angles for the two joints of the shoulder, upper arm and elbow joints. In this simulation, objects are not present.

When performing the action, the teacher agent can emit a three-letter word labeling the action. Each label consists of a Consonant–Vowel–Consonant word, such as "XUM", "HAW", "RIV". All consonants and letters of the English alphabet are used. Each letter is encoded using two real-value features in the interval [0,1]. Therefore each action word is represented by 6 features. Each word is unique to the action performed.

B. MFT algorithm

We use a multi-dimensional MFT algorithm (Tikhanoff et al. 2006) with 112 input fields (concept-models) randomly initialized. We consider the action learning problem as that of categorizing N=112 objects (actions) i=1,…,N, each of which is characterized by d=12 features e=1,…,d. These features are represented by real numbers $O_{ie} \in (0,1)$ – the input signals. These 12 features correspond to the 6 joint rotation angles and 6 phonetic encoding values. Moreover, we assume that there are M=112 d-dimensional fields (i.e. concept-models of the prototype of actions/words to be learned) k=1,…,M described by real-valued fields S_{ke}, with e=1,…,d as before. The concept models will tend to match the input object features O_{ie} during learning by maximizing the global similarity function

$$L = \sum_i \log \sum_k l(i \mid k)$$

where

$$l(i \mid k) = \prod_{e=1}^{d} \left(2\pi\sigma_{ke}^2\right)^{-1/2} \exp\left[-\left(S_{ke} - O_{ie}\right)^2 / 2\sigma_{ke}^2\right]$$

is the similarity measure between object i and concept k. Here σ_{ke} is the fuzziness parameter that gradually decreases over time. Full details on the learning algorithm can be found in Tikhanoff et al. 2007. See also Perlovsky (this volume) for an overview of the MFT algorithm.

C. Results

The simulation lasts for 25000 training steps. In the first 12500 cycles, only the 6 action features (angles) are provided. This is enough for the agents to learn to reproduce the action repertoire. At cycle 12500 (half of the training time), all 12 feature sets (6 for actions/angles, 6 for phonetic sounds) are considered when computing the MFT fuzzy similarity functions. The re-initialization of the fuzziness parameter σ_{ke} at timestep 125000 allows the agent to learn the new sound features and create a concept model of the labels.

Results demonstrate that the robot is able to categorize 95% of actions and learn their unique labels. Figure 3 shows the evolution of the 112 concept-model fields during training. Note the resetting of the fields at timestep 12500, when words are introduced and the fuzziness σ_{ke} is reinitialized.

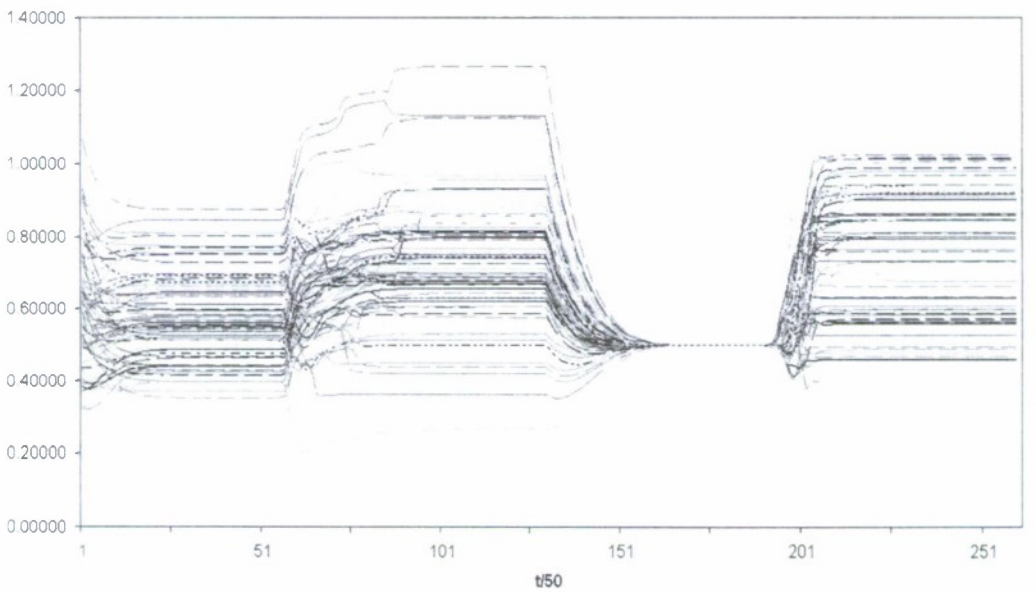

Figure 3 – Evolution of 112 concept-models during training. Vertical axis represents a compressed one-dimensional representation of the concept-models using the amplitude $S_k = \sqrt{\sum_{e=1}^{d} (S_{ke})^2} \Big/ d$.

IV. Discussion and Conclusions

The simulation experiments above concern the study of the language grounding in action and the symbol grounding transfer in cognitive robotic agents. The positive results of the grounding transfer simulation demonstrate that it is possible to design autonomous linguistic agents capable of acquiring new grounded concepts.

The use of MFT to overcome the CC limitations of connectionist models demonstrates that it is possible to scale up the action and lexicon repertoire of the cognitive robotic agents. Perlovsky's (2004)

proposal to apply MFT in the language domain is highly consistent with the grounded approach to language modeling discussed above. That is, both accounts are based on the strict integration of language and cognition. This permits the design of cognitive systems that are truly able to "understand" the meaning of words being used by autonomously linking the linguistic signals to the internal concept-models of the word constructed during the sensorimotor interaction with the environment.

The potential impact of this grounded cognitive robotic approach for the development of intelligent systems is great, both for cognitive science and for technology. In cognitive science, the area of embodied cognition regards the study of the functioning and organization of cognition in natural and artificial systems. For example, the Higher-Order learning procedure is inspired by Barsalou's "reenactment" and "mental simulation" mechanism in the perceptual symbol system hypothesis. Barsalou (1997) demonstrates that during perceptual experience, association areas in the brain capture bottom-up patterns of activation in sensory-motor areas. Later, in a top-down manner, association areas partially reactivate sensory-motor areas to implement perceptual symbols simulators. A simulation platform like the one used here can be used to test further embodied cognition theories of language, such as Glenberg and Kaschak's (2002) action-compatibility effects. In addition, such an approach can be used to study the development and emergence of language in epigenetic robots (Weng et al. 2001; Metta et al. 2006)

For the technological implications of such a project, the model proposed here can be useful in fields such as that of defense systems, service robotics and human-robot interaction. In the area of defense systems, cognitive systems are essential for integrated multi-platform systems capable of sensing and communicating. Such robots can be beneficial in collaborative and distributed tasks such as multi-agent exploration and navigation in unknown terrains. In service and household robotics, future systems will be able to learn language and world understanding from humans, and also to interact with them for entertainment purposes (e.g. Tikhanoff & Miranda, 2005; Steels & Kaplan 2000). In human-robot communication systems, robots will develop their lexicon through close interaction with their environment and whilst communicating with humans. Such a social learning context can permit a more efficient acquisition of communication capabilities in autonomous robots, as demonstrated in Steels & Kaplan (2000).

REFERENCES

Barsalou L. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, 22, 577-609.

Cangelosi (2005). Approaches to Grounding Symbols in Perceptual and Sensorimotor Categories. In H. Cohen & C. Lefebvre (Eds), *Handbook of Categorization in Cognitive Science*, Elsevier, pp. 719-737

Cangelosi A. & Harnad S. (2000). The adaptive advantage of symbolic theft over sensorimotor toil: Grounding language in perceptual categories. *Evolution of Communication* 4(1), 117-142

Cangelosi A., G. Bugmann and R. Borisyuk (Eds.), *Modeling Language, Cognition and Action: Proceedings of the 9th Neural Computation and Psychology Workshop*. Singapore: World Scientific, 2005

Cangelosi A., Greco A. & Harnad S. (2000). From robotic toil to symbolic theft: Grounding transfer from entry-level to higher-level categories. Connection Science, 12(2), 143-162

Cangelosi A., Hourdakis E. & Tikhanoff V. (2006). Language acquisition and symbol grounding transfer with neural networks and cognitive robots. *Proceedings of IJCNN 2006*, Vancouver

Cangelosi, A. and T. Riga, "An embodied model for sensorimotor grounding and grounding transfer: Experiments with epigenetic robots". *Cognitive science*, in press2006, 30(4), 673-689

Feldman, J., & Narayanan S. (2004). Embodied meaning in a neural theory of language. *Brain and Language*, 89, 385-392.

Glenberg A., and K. Kaschak, "Grounding language in action," *Psychonomic Bulletin & Review*, vol. 9(3), pp. 558-565, 2002.

Harnad, S. "The Symbol Grounding Problem," *Physica D*, vol. 42, pp. 335-346, 1990.

L. Perlovsky, L. "Integrating language and cognition," *IEEE Connections*, vol. 2(2), pp. 8-13, 2004

Metta G., Fitzpatrick P., Natale L. (2006). YARP: yet another robot platform. *International Journal on Advanced Robotics Systems*, 3, 43-48.

Nolfi S. & Floreano D. (2000). *Evolutionary Robotics: The Biology, Intelligence, and Technology of Self-Organizing Machines*. Cambridge, MA: MIT Press/Bradford Books.

Pecher, D., & Zwaan, R.A., (Eds.). (2005). *Grounding cognition: The role of perception and action in memory, language, and thinking*. Cambridge, UK: Cambridge University Press.

Perlovsky L. (2001). *Neural Networks and Intellect: Using Model-Based Concepts*. Oxford University Press, New York.

Plunkett K., Sinha C., Møller M.F., Strandsby O. (1992). Symbol grounding or the emergence of symbols? Vocabulary growth in children and a connectionist net. *Connection Science*, 4: 293-312

Pulvermuller F. (2003). *The Neuroscience of Language. On Brain Circuits of Words and Serial Order*. Cambridge University Press.

Rizzolatti G. & Arbib M.A. (1998). Language within our grasp. *Trends in Neurosciences*, 21(5), 188-194.

Steels L., and K. Kaplan, "AIBO's first words: The social learning of language and meaning," *Evolution of Communication*, vol. 4(1), pp. 3-32, 2000.

Talmy L. (2000). *Toward a Cognitive Semantics*, Vol. I. Cambridge: Cambridge University Press

Tikhanoff V., and E.R. Miranda, "Musical Composition by an Autonomous Robot: An Approach to AIBO Interaction", in *Proceedings of TAROS 2005 (Towards Autonomous Robotic System)*, Imperial College, London, UK. 2005, pp. 181-188.

Tikhanoff V., Cangelosi A., Fontanari J.F. & Perlovsky L.I (2007). Scaling up of action repertoire in linguistic cognitive agents. *Proceedings IEEE-KIMAS'07 Conference*. Waltham, MA

Tikhanoff V., Fontanari J.F., Cangelosi A. & Perlovsky L.I. (2006). Language and cognition integration through modeling field theory: Category formation for symbol grounding. *ICANN06 International Conference on Artificial Neural Networks*, Athens, September 2006.

Weng J., McClelland J., Pentland A., Sporns O., Stockman I., Sur M. & Thelen E. (2001). Autonomous mental development by robots and animals. *Science*, 291, 599-600.

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin and Review*, 9, 625-636.